

Baifeng Shi

baifeng_shi@berkeley.edu | (+1) 510-495-7418

<https://bfshi.github.io/>

EDUCATION BACKGROUND

University of California, Berkeley	09/2021 – 06/2026 (expected)
<i>Ph.D. student, Computer Science</i>	
• Advisor: Trevor Darrell	
• Awards and Honors:	
BAIR Ignition Reward, UC Berkeley	09/2021
Peking University	09/2017 - 06/2021
<i>B.S., Computer Science</i>	
• Advisor: Yadong Mu	
• Overall GPA: 3.75 / 4	
• Awards and Honors:	
Gold Medal (3 / 360), Chinese Physics Olympiad final contest	10/2016
EECS Dean Scholarship, Peking University	09/2017
Merit Student, Peking University	09/2018 & 09/2020

RESEARCH APPOINTMENTS

• University of California, Berkeley , Research Assistant	09/2021 - now
<i>Advisor: Prof. Trevor Darrell</i>	
• Microsoft Research , Research Collaboration through BAIR Commons Program	09/2021 - 02/2022
<i>Advisor: Dr. Xin Wang & Prof. Trevor Darrell</i>	
• University of California, Berkeley , Research Intern	03/2020 - 11/2020
<i>Advisor: Dr. Huijuan Xu & Prof. Trevor Darrell</i>	
• Microsoft Research Asia , Research Intern	09/2019 - 03/2021
<i>Advisor: Dr. Qi Dai & Dr. Jingdong Wang</i>	
• Peking University , Research Intern	09/2018 - 09/2019
<i>Advisor: Prof. Yadong Mu</i>	

PUBLICATIONS

-
- Ilija Radosavovic, **Baifeng Shi**, Letian Fu, Ken Goldberg, Trevor Darrell*, Jitendra Malik*, *Robot Learning with Sensorimotor Pre-training*, **submitted to CoRL 2023**
 - **Baifeng Shi**, Siyu Gai, Trevor Darrell, Xin Wang, *TOAST: Transfer Learning via Attention Steering*, **submitted to NeurIPS 2023**
 - **Baifeng Shi**, Trevor Darrell, Xin Wang, *Top-down Visual Attention from Analysis by Synthesis*, **CVPR 2023, conference highlight**
 - **Baifeng Shi**, Yale Song, Neel Joshi, Trevor Darrell, Xin Wang, *Visual Attention Emerges from Recurrent Sparse Reconstruction*, **ICML 2022**
 - **Baifeng Shi**, Qi Dai, Judy Hoffman, Kate Saenko, Trevor Darrell, Huijuan Xu, *Temporal Action Detection with Multi-level Supervision*, **ICCV 2021**
 - **Baifeng Shi**, Judy Hoffman, Kate Saenko, Trevor Darrell, Huijuan Xu, *Auxiliary Task Reweighting for Minimum-data Learning*, **NeurIPS 2020**
 - Zhekun Luo, Devin Guillory, **Baifeng Shi**, Wei Ke, Fang Wan, Trevor Darrell, Huijuan Xu, *Weakly-Supervised Action Localization with Expectation-Maximization Multi-Instance Learning*, **ECCV 2020**
 - **Baifeng Shi***, Dinghuai Zhang*, Qi Dai, Zhanxing Zhu, Yadong Mu, Jingdong Wang, *Informative Dropout for Robust Representation Learning: A Shape-bias Perspective*, **ICML 2020**
 - **Baifeng Shi**, Qi Dai, Jingdong Wang, Yadong Mu, *Weakly-Supervised Action Localization by Generative Attention Modeling*, **CVPR 2020**

RESEARCH EXPERIENCE

Independent Research, advised by Dr. Xin Wang and Prof. Trevor Darrell

UC Berkeley

TOAST: Transfer Learning via Attention Steering

02/2022 – 06/2022

- Showed that current transfer learning methods (*e.g.*, fine-tuning, LoRA, prompt tuning) often fail to focus the model's attention on features relevant to downstream tasks.
- Proposed TOAST, which refocuses the attention using our top-down attention algorithm and showed a significant improvement over previous baselines on transfer learning performances for visual fine-grained classification, .
- Also applied TOAST to instruction-following language generation and showed improvement over fine-tuned Alpaca and Vicuna models.

Independent Research, advised by Dr. Xin Wang and Prof. Trevor Darrell

UC Berkeley

Top-down Visual Attention from Analysis by Synthesis

03/2022 – 02/2023

- Current self-attention is bottom-up, highlighting all the objects in an image and mix the representations of all objects together. In contrast, top-down attention enables us to only attend to the objects that are related to the current task or goal.
- Showed that by formulating the visual recognition process as Bayesian Inference (Analysis by Synthesis), top-down attention will naturally emerge, where the high-level goal acts as a prior to modulate the low-level attention.
- Proposed AbSViT, a ViT model with a feedback loop which is trained to variationally approximate Analysis by Synthesis. Showed that AbSViT learns to control the top-down attention given language or class guidance.
- Applied AbSViT in various real-world scenarios, such as vision-language tasks of VQA and zero-shot recognition where a language prompt guides the attention, as well as image classification and robustness where no specific guidance is available.

Independent research, advised by Dr. Xin Wang and Prof. Trevor Darrell

UC Berkeley, Microsoft Research

Visual Attention Emerges from Recurrent Sparse Reconstruction

08/2021 – 02/2022

- Proposed a formulation of visual attention by taking inspiration from how human visual attention works, especially focusing on two key ingredients in human visual attention, recurrent connection and sparse representation
- Showed that adding recurrent connections into feedforward network is equivalent to adding sparse reconstruction blocks.
- Showed that attention can naturally emerges from sparse reconstruction, where recurrent connections group features into separate objects and the sparsity constraint selects the most salient objects.
- Pointed out that self-attention is a special case of our attention formulation, providing a new perspective on how self-attention works.
- Observed higher robustness and more consistency with human eye fixation map.
- The work is summarized in a paper and accepted to ICML 2022.

Independent research, advised by Dr. Huijuan Xu and Prof. Trevor Darrell

UC Berkeley

Unsupervised Foreground Mining for Omni-supervised Action Localization

06/2020 – 11/2020

- Built the first baselines for semi-supervised and omni-supervised action localization.
- Designed error analysis to find the main sources of error in the baseline models.
- Proposed to solve the action incompleteness problem in the semi-supervised baseline by learning object-centric representations. Built a structural causal model of the foreground/background action in neighboring frames, and proposed to detect foreground objects by minimizing the conditional mutual information between foreground and background motion.
- Proposed to solve the action-context confusion problem in the omni-supervised baseline by designing an information bottleneck to discard scene information while preserve action information.
- The work is summarized in a paper and accepted to ICCV 2021.

Independent research, advised by Dr. Huijuan Xu and Prof. Trevor Darrell

UC Berkeley

Auxiliary Task Reweighting for Minimum-data Learning

03/2020 – 06/2020

- Addressed the problem of automatically reweighting multiple auxiliary tasks to learn the main task with minimum information (supervision).
- Exploited the key insight that *information required for inference can be reduced by a good prior*, and formulated the problem as optimizing the KL divergence between the true prior and the surrogate prior given by the weighted likelihood of auxiliary tasks.
- Utilized tools and concepts including Fisher score and Langevin dynamics, and further simplified the optimization of the KL divergence into minimizing the l2 distance between main/auxiliary task gradients, which gives a light-weight algorithm to reweight auxiliary tasks on-the-fly.

- Derived theoretical guarantees that our algorithm finds the optimal task weights up to a small error.
- Experimentally observed that our algorithm finds the optimal task weights and minimizes the data requirement under various settings, *e.g.*, semi-supervised learning, domain generalization, and multi-label classification.
- The work is summarized in a paper and accepted to NeurIPS 2020.

Independent research, mentored by Dr. Qi Dai, Dr. Jingdong Wang, and Prof. Yadong Mu *Microsoft Research Asia*
Human Vision Inspired Shape-bias for Model Robustness 11/2019 – 02/2020

- Analyzed the relationship between the texture-bias of CNN and its multiple kinds of vulnerability.
- Proposed to discriminate texture from shape by the self-information in an image, resembling the mechanism of saliency detection and eye movement in the human visual system.
- Proposed a Dropout-like algorithm to de-correlate the model output with the texture information in the input, thus enhancing the shape-bias of the model.
- Conducted experiments under different scenarios (domain generalization, few-shot classification, image corruption, and adversarial perturbation) and observed a universal improvement in model robustness.
- The work is summarized in a paper and accepted to ICML 2020.

Independent research, mentored by Dr. Qi Dai, Dr. Jingdong Wang, and Prof. Yadong Mu *Microsoft Research Asia*
Weakly Supervised Action Localization 08/2019 – 11/2019

- Tackled with a common challenge in weakly supervised action localization, namely action-context confusion.
- Built a probabilistic graphical model and formulated the problem as modeling the frame-wise class-agnostic likelihood and optimizing the maximum a posteriori (MAP) estimation.
- Proposed to separate foreground and context by modeling the appearance-level frame likelihood using a generative model, *viz.* conditional variational auto-encoder (CVAE).
- Improved the results on two common datasets by a large margin (10% relative improvement).
- The work is summarized in a paper and accepted to CVPR 2020.

Independent research, advised by Prof. Yadong Mu *Peking University*
Fast Video Understanding with Reinforcement Learning 10/2018 – 08/2019

- Proposed an algorithm to use as few frames as possible to classify a video while preserving an expected accuracy.
- Formulated the reward function as the Lagrangian of the constrained optimization problem, and optimized it using reinforcement learning algorithm, *viz.* soft actor-critic (SAC).
- Boost the original algorithm by 400% while preserving the accuracy.