Amanpreet Singh

https://apsdehal.in
me@apsdehal.in
March 2022

Research Interests	Multimodal, Computer Vision, Natural Language Processing, Multiagent Systems

Professional Experience

HuggingFace, *Lead Researcher - Palo Alto, CA*   March 2022 2018 - Current

- Working on cutting-edge multimodal research with emphasis on open science.

Facebook, *Research Engineer - Menlo Park, CA*   July 2018 - March 2022

- Designed and developed a framework (MMF) for multimodal research and production use cases. Used heavily internally (33 teams) as well as externally.
- Led and executed different research directions with emphasis on large scale vision and language applications.

Amazon, *Software Developer Intern - Seattle*   May 2017 - Aug 2017

- Designed and developed a service for detecting compatibility issues between tightly-coupled services.
- Implemented the solution in a generic way as issue was widespread due to rapid development of services.

The Climate Corporation, *Software Intern - SFO*   May 2015 - July 2015

- Developed a production Clojure service for spatially searching over 1B US fields and processing complex spatial queries in seconds rather than hours.
- Built a ClojureScript tracker service to measure and maintain internal services.

Education

New York University (NYU)	Sep 2016 - May 2018
Masters in Computer Science	GPA: 4.0/4.0

Indian Institute of Technology, Roorkee	Aug 2012 - May 2016
Bachelors in Electrical Engineering	

Publications & Preprints

[1] Mingyang Zhou, Licheng Yu, **Amanpreet Singh**, Mengjiao Wang, Zhou Yu, and Ning Zhang. Unsupervised vision-and-language pre-training via retrieval-based multi-granular alignment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022.

[2] **Amanpreet Singh**, Ronghang Hu, Vedanuj Goswami, Guillaume Couairon, Wojciech Galuba, Marcus Rohrbach, and Douwe Kiela. Flava: A foundational language and vision alignment model. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022.

[3] Sasha Sheng, **Amanpreet Singh**, Vedanuj Goswami, Jose Magana, Tristan Thrush, Wojciech Galuba, Devi Parikh, and Douwe Kiela. Human-adversarial

visual question answering. *Advances in Neural Information Processing Systems*, 34, 2021.

[4] Douwe Kiela, Max Bartolo, Yixin Nie, Divyansh Kaushik, Atticus Geiger, Zhengxuan Wu, Bertie Vidgen, Grusha Prasad, **Amanpreet Singh**, Pratik Ringshia, et al. Dynabench: Rethinking benchmarking in nlp. In *NAACL-HLT*, 2021.

[5] Ronghang Hu and **Amanpreet Singh**. Unit: Multimodal multitask learning with a unified transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1439–1449, 2021.

[6] **Amanpreet Singh**, Guan Pang, Mandy Toh, Jing Huang, Wojciech Galuba, and Tal Hassner. Textocr: Towards large-scale end-to-end reasoning for arbitrary-shaped scene text. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8802–8812, 2021.

[7] Medhini Narasimhan, Erik Wijmans, Xinlei Chen, Trevor Darrell, Dhruv Batra, Devi Parikh, and **Amanpreet Singh**. Seeing the un-scene: Learning amodal semantic maps for room navigation. In *Proceedings of IEEE/CVF European Conference on Computer Vision (ECCV)*, 2020.

[8] Oleksii Sidorov, Ronghang Hu, Marcus Rohrbach, and **Amanpreet Singh**. Textcaps: a dataset for image captioning with reading comprehension. In *Proceedings of IEEE/CVF European Conference on Computer Vision (ECCV)*, 2020.

[9] **Amanpreet Singh**, Vedanuj Goswami, and Devi Parikh. Are we pretraining it right? digging deeper into visio-linguistic pretraining. *arXiv preprint arXiv:2004.08744*, 2020.

[10] Douwe Kiela, Hamed Firooz, Aravind Mohan, Vedanuj Goswami, **Amanpreet Singh**, Pratik Ringshia, and Davide Testuggine. The hateful memes challenge: Detecting hate speech in multimodal memes. *Advances in Neural Information Processing Systems*, 33:2611–2624, 2020.

[11] Ronghang Hu, **Amanpreet Singh**, Trevor Darrell, and Marcus Rohrbach. Iterative answer prediction with pointer-augmented multimodal transformers for textvqa. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9992–10002, 2020.

[12] Alex Wang*, Yada Pruksachatkun*, Nikita Nangia*, **Amanpreet Singh***, Julian Michael, Felix Hill, Omer Levy, and Samuel Bowman. Superglue: A stickier benchmark for general-purpose language understanding systems. In *Advances in Neural Information Processing Systems*, pages 3266–3280, 2019.

[13] **Amanpreet Singh**, Vivek Natarajan, Meet Shah, Yu Jiang, Xinlei Chen, Dhruv Batra, Devi Parikh, and Marcus Rohrbach. Towards vqa models that can read. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8317–8326, 2019.

[14] **Amanpreet Singh**, Tushar Jain, and Sainbayar Sukhbaatar. Learning when to communicate at scale in multiagent cooperative and competitive tasks. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2019.

[15] Nan Rosemary Ke, **Amanpreet Singh**, Ahmed Touati, Anirudh Goyal, Yoshua Bengio, Devi Parikh, and Dhruv Batra. Learning dynamics model in reinforcement learning by incorporating the long term future. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2019.

[16] Alex Wang, **Amanpreet, Singh**, Julian Michael, Felix Hill, Omer Levy, and Samuel R Bowman. Glue: A multi-task benchmark and analysis platform for natural language understanding. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2019.

[17] Alex Warstadt, **Amanpreet Singh**, and Samuel R Bowman. Neural network acceptability judgments. *Transactions of the Association for Computational Linguistics*, 7:625–641, 2019.

[18] **Amanpreet Singh**, Vivek Natarajan, Yu Jiang, Xinlei Chen, Meet Shah, Marcus Rohrbach, Dhruv Batra, and Devi Parikh. Mmf - a multimodal framework for vision & language research. In *SysML Workshop, NeurIPS*, volume 2018, 2018.

[19] **Amanpreet Singh** and Sharan Agrawal. Canvasgan: A simple baseline for text to image generation by incrementally patching a canvas. In *Science and Information Conference*, pages 86–98. Springer, 2019.

Honor & Awards	**Oral Presentation, TextCaps (2.1% papers)** (ECCV, 2020)	2019
	Oral Presentation, M4C (2.3% papers) (CVPR, 2019)	2019
	Spotlight Presentation, SuperGLUE (4.2% papers) (NeurIPS, 2019)	2019
	Fellowship for best research project, NYU (Masters Scholarship)	2018
	Fellowship for best innovation project, NYU (Masters Scholarship)	2017
	Fellowship for top student in state, IITR (Undergraduate Scholarship)	2012
	3rd Place, TopCoder Hackathon (San Francisco)	2015
Invited Talks	*Towards unified foundation model for vision-and-language alignment* Microsoft Research	Mar 2022
	Towards unified foundation model for vision-and-language alignment MLRW, IITG	Feb 2022
	Towards models that can read and reason about scene text DocVQA Workshop, KDD 2021	Sep 2021
	TextCaps dataset and challege 2021 Visual QA and Dialog Workshop, CVPR 2021	Jun 2021
	Building knowledge based VQA systems with MMF KDD 2020	Oct 2020
	Supercharing Vision and Language Reasearch with MMF Visual QA and Dialog Workshop, CVPR 2020	Jun 2020
	TextVQA dataset and challenge 2020 Visual QA and Dialog Workshop, CVPR 2020	Jun 2020
	TextVQA dataset and challenge 2019 Visual QA and Dialog Workshop, CVPR 2019	Jun 2020

Professional Activities

Reviewer: CVPR 2022, ICLR 2022, ICCV 2021, NeurIPS 2021, ECCV 2020, CVPR 2021, ICLR 2021, CVPR 2020, EMNLP 2020, ACL 2020, ICLR 2020, NeurIPS 2020, AISTATS 2020, MVAP Journal, TPAMI, and others

Workshop Organization: Visual Question Answering and Dialog Workshop, CVPR 2020; Visual Question Answering and Dialog Workshop, CVPR 2019.

Panelist: Visual Question Answering and Dialog Workshop, CVPR 2020.

Tutorials: Building knowledge based VQA systems with MMF, KDD 2020.

Organization: jQuery Foundation, PyTorch