HiKER-SGG: Hierarchical Knowledge Enhanced Robust Scene Graph Generation

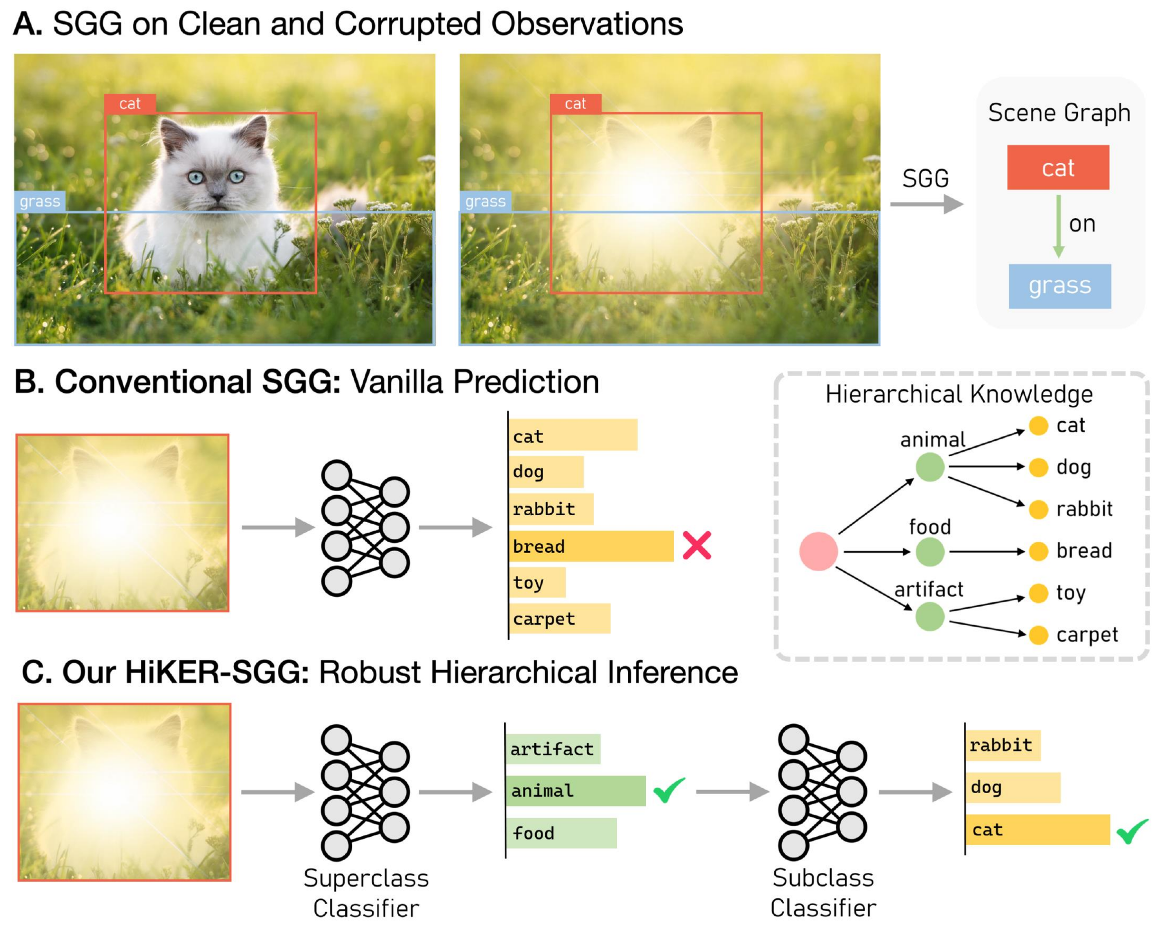
Ce Zhang Simon Stepputtis Joseph Campbell Katia Sycara Yaqi Xie

Carnegie Mellon University School of Computer Science · The Robotics Institute · AFRL Air Force Office of Scientific Research · DEVCOM Army Research Laboratory

CVPR JUNE 17-21, 2024 SEATTLE, WA

Motivation for *Robust* Scene Graph Generation

- **Scene Graph Generation (SGG)** from visual inputs is a powerful method of extracting semantic information, enabling many subsequent reasoning tasks.
- However, most existing studies assume access to **"clean" images**. This contrasts with real-world situations where images often have corruptions like sun glare, dust, *etc*.
- Handling such corruptions is a challenging task as it is **unlikely that models can be sufficiently trained** to handle such domain shifts.

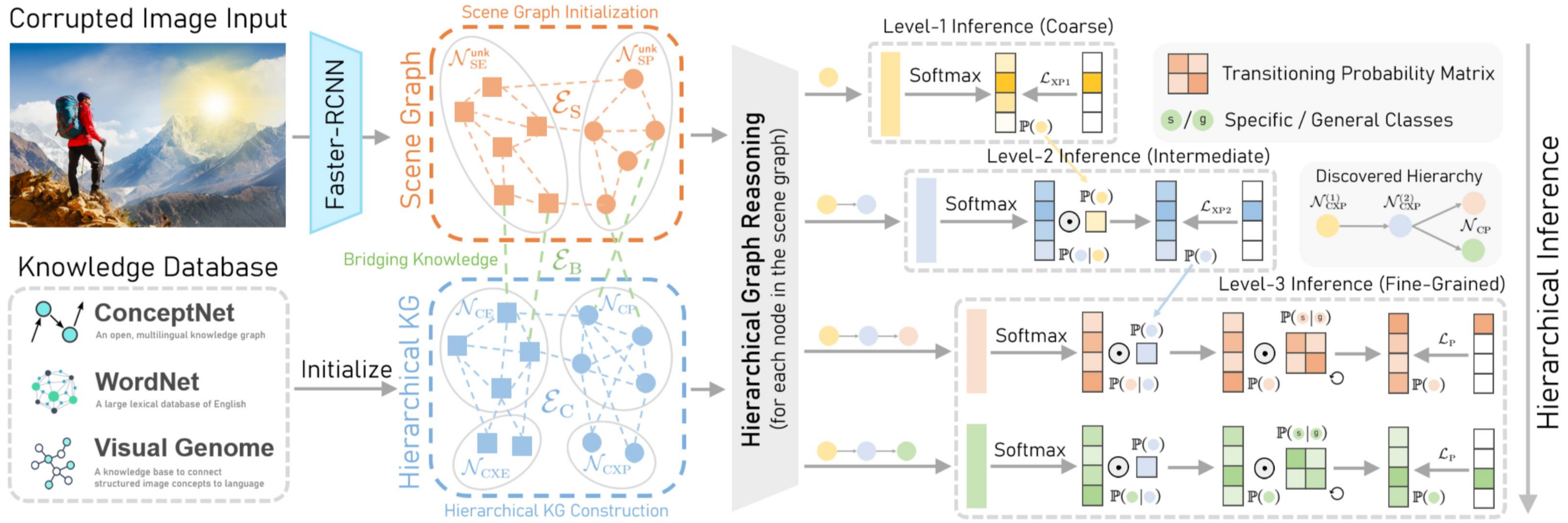


Corrupted Visual Genome (VG-C) Benchmark

We introduce a challenging benchmark – Corrupted Visual Genome (VG-C) – providing **20 procedurally generated image corruptions**, including simple transformations and severe weather conditions. VG-C benchmark offers a comprehensive evaluation platform to **assess the robustness of SGG models** in adverse conditions.

HiKER-SGG Framework Highlights

HiKER-SGG: Hierarchical **K**nowledge **E**nhanced **R**obust **S**cene **G**raph **G**eneration



★ **Overview.** In this work, we propose a novel method – HiKER-SGG – which utilizes a hierarchical approach that reasons over *multiple levels of domain knowledge* with *increasing granularity* in order to generate accurate scene graphs in adverse conditions.

★ **Hierarchical Knowledge.** The hierarchical knowledge graph is constructed from *external knowledge base*. We define it as comprising a set of commonsense *object* and *relation* nodes, and *superclass* nodes. The representations of superclass nodes are initialized by averaging the children nodes representations.

★ **Hierarchical Inference.** HiKER-SGG employs a *hierarchical* strategy, leveraging multi-level domain knowledge to produce precise scene graphs for *both corrupted and clean* images. It classifies objects/relations by *initially determining their superclasses*, subsequently *narrowing the search space to the respective subclasses*. This method simplifies the process compared to the conventional direct detailed prediction.

★ **Adaptive Refinement.** HiKER-SGG integrates an adaptive refinement mechanism to mitigate biases in predicate classes. Specifically, our adaptive refinement dynamically updates the transition probabilities that can convert a *general* prediction into a more *specific* prediction during the training process.

★ **Corruption Agnostic.** Note that all models *are solely trained on clean* images and *directly tested on corrupted* ones.

Takeaways

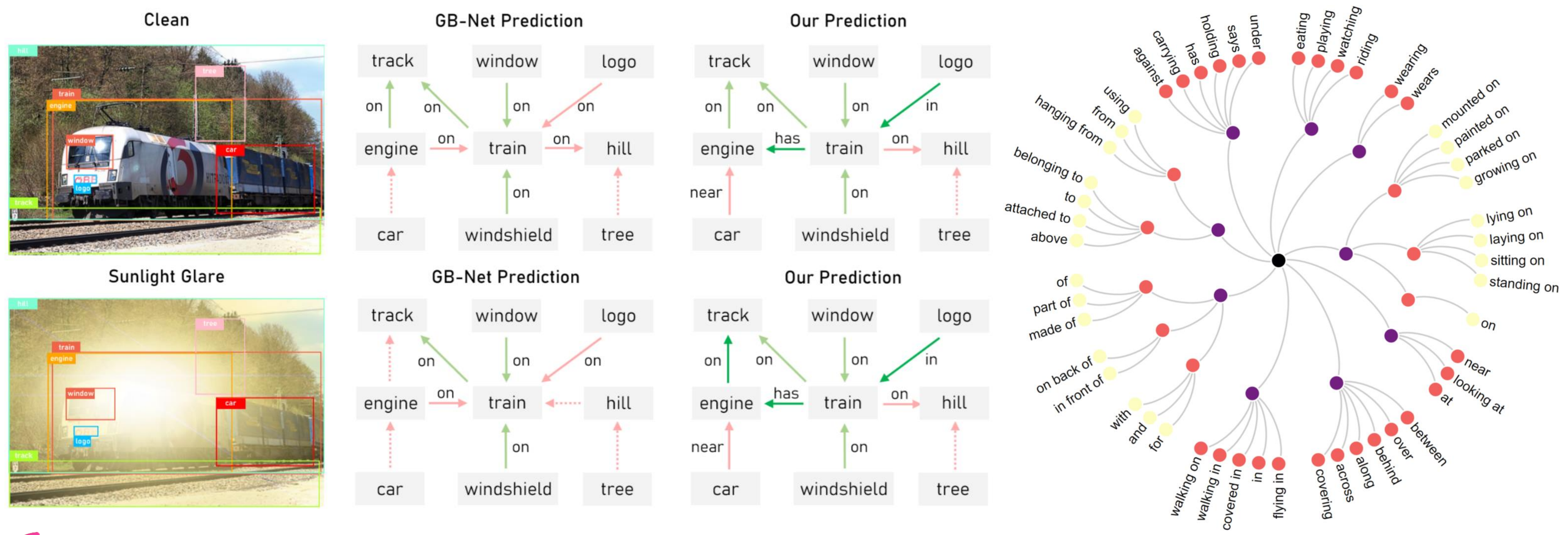
✓ We propose HiKER-SGG, a novel method for generating scene graphs via **hierarchical inference** over structured domain knowledge, allowing it to gradually specify **increasingly granular classifications** through **iterative sub-selection**.

✓ We introduce a new synthetic VG-C benchmark for SGG, containing **20 challenging image corruptions**, including simple transformations and severe weather conditions.

✓ Extensive experiments demonstrate that HiKER-SGG **outperforms current state-of-the-art methods** on SGG tasks, while simultaneously providing **a strong zero-shot baseline** for generating scene graphs from **corrupted images**.

Experiments

Qualitative Results



Evaluation on **Clean** Visual Genome (VG) Dataset

Method	Venue	PredCls mR@20: UC/C	PredCls mR@50: UC/C	PredCls mR@100: UC/C	SGCls mR@20: UC/C	SGCls mR@50: UC/C	SGCls mR@100: UC/C
IMP+ [75]	*CVPR'17*	- / -	20.3 / 9.8	28.9 / 10.5	- / -	12.1 / 9.8	16.9 / 10.5
Neural Motifs [87]	*CVPR'18*	- / 10.8	24.8 / 14.0	37.3 / 15.3	- / 6.3	13.5 / 7.7	19.6 / 8.2
VCTree [62]	*CVPR'19*	- / 14.0	- / 17.9	- / 19.4	- / 8.2	- / 10.1	- / 10.8
PCPL [77]	*ACMMM'20*	- / -	50.6 / 35.2	62.6 / 37.8	- / -	<u>26.8</u> / 18.6	<u>32.8</u> / 19.6
Transformer + CogTree [84]	*IJCAI'21*	- / 22.9	- / 28.4	- / 31.0	- / 13.0	- / 15.7	- / 16.7
VCTree + EBM [60]	*CVPR'21*	- / 14.2	- / 18.0	- / 28.8	- / 8.2	- / 10.2	- / 11.0
G2S: Transformer [22]	*ICCV'21*	- / 26.7	- / 31.9	- / 34.2	- / 15.7	- / 18.5	- / 19.4
MotifNet + DLFE [10]	*ACMMM'21*	- / 22.1	- / 26.9	- / 28.8	- / 12.8	- / 15.2	- / 15.9
MotifNet + RTPB [6]	*AAAI'22*	- / 28.8	- / 35.3	- / 37.7	- / 16.3	- / 19.4	- / 20.6
MotifNet + PPDL [43]	*CVPR'22*	- / 27.9	- / 32.2	- / 33.3	- / 15.8	- / 17.5	- / 18.2
MotifNet + NICE [40]	*CVPR'22*	- / 23.7	- / 29.8	- / 32.2	- / 13.6	- / 16.7	- / 17.9
MotifNet + NARE [19]	*CVPR'22*	- / 21.3	- / 27.1	- / 29.7	- / 11.3	- / 14.3	- / 15.7
Transformer + HML [13]	*ECCV'22*	- / 27.4	- / 33.3	- / 35.9	- / 15.7	- / 19.1	- / 20.4
SQUAT [36]	*CVPR'23*	- / 25.6	- / 30.9	- / 33.4	- / 14.4	- / 17.5	- / 18.8
PE-Net [92]	*CVPR'23*	- / 25.8	- / 31.4	- / 33.5	- / 15.2	- / 18.2	- / 19.3
PE-Net + SIL [69]	*ACMMM'23*	- / 26.9	- / 33.1	- / 35.3	- / <u>16.7</u>	- / <u>19.9</u>	- / <u>20.7</u>
GB-Net [85]	*ECCV'20*	23.8 / 15.3	41.1 / 19.3	55.4 / 20.9	13.1 / 7.9	21.4 / 9.6	29.1 / 10.2
EB-Net + EOA [9]	*WACV'23*	<u>39.8</u> / <u>30.8</u>	<u>54.9</u> / <u>36.7</u>	<u>66.3</u> / <u>39.2</u>	<u>19.6</u> / 14.9	26.7 / 17.3	32.5 / 18.3
HiKER-SGG (Ours)	-	**42.1 / 33.4**	**57.9 / 39.3**	**69.2 / 41.2**	**22.6 / 18.2**	**30.0 / 20.3**	**36.7 / 21.4**

Evaluation on **Corrupted** Visual Genome (VG-C) Dataset

	Method	gaus	shot	imp	dfcs	gls	mtn	zm	snw	frst	fg	brt	cnt	els	px	jpg	sun	wtd	smk	rain	dust	Average mR
mR@20: C/UC	GB-Net† [85]	15.2	16.0	15.2	16.9	14.9	16.5	16.6	17.9	18.9	21.4	21.6	14.7	16.8	16.6	18.2	16.7	17.8	16.0	20.1	18.5	17.3 (-27.3%)
	EB-Net† [9]	28.0	29.8	27.4	31.2	26.5	30.3	30.5	32.1	33.2	35.8	36.3	27.3	30.3	27.0	30.6	30.6	30.7	33.7	35.6	30.1	30.9 (-22.4%)
	HiKER-SGG	**31.1**	**33.3**	**31.5**	**35.4**	**28.5**	**35.0**	**34.1**	**36.5**	**37.7**	**39.8**	**40.8**	**30.5**	**33.7**	**31.3**	**34.2**	**33.5**	**34.9**	**37.1**	**39.8**	**32.6**	**34.6 (-17.8%)**
	GB-Net† [85]	10.3	10.6	10.4	11.6	10.4	10.9	10.7	11.9	12.3	13.7	13.8	10.0	11.1	10.8	11.7	11.1	11.2	10.5	13.0	12.1	11.4 (-25.5%)
	EB-Net† [9]	21.7	22.8	20.4	24.9	19.6	23.2	23.8	23.2	24.6	27.5	28.0	20.1	23.1	21.1	23.6	24.0	23.4	25.6	27.3	22.9	23.5 (-23.7%)
	HiKER-SGG	**24.8**	**25.8**	**24.8**	**27.5**	**22.4**	**27.4**	**26.4**	**27.8**	**28.7**	**31.1**	**31.5**	**23.3**	**26.0**	**24.3**	**26.5**	**26.3**	**26.8**	**28.5**	**30.9**	**24.9**	**26.8 (-19.8%)**
mR@50: C/UC	GB-Net† [85]	27.5	28.7	27.6	30.8	26.4	29.8	29.9	31.9	33.8	37.2	37.6	26.3	29.9	30.0	33.0	29.5	32.3	28.7	35.8	32.8	31.0 (-24.6%)
	EB-Net† [9]	42.1	43.7	41.5	44.9	40.2	45.6	44.2	46.9	47.7	50.4	51.2	41.2	44.1	41.4	45.1	45.4	45.5	48.4	49.7	44.6	45.2 (-17.7%)
	HiKER-SGG	**46.7**	**48.4**	**46.9**	**50.2**	**43.2**	**49.6**	**48.3**	**51.3**	**52.5**	**55.1**	**55.9**	**45.0**	**48.1**	**46.0**	**49.9**	**48.6**	**50.0**	**52.4**	**54.8**	**47.0**	**49.5 (-14.5%)**
	GB-Net† [85]	13.3	13.6	13.3	15.1	13.6	14.1	14.0	15.4	15.6	17.4	17.5	13.0	14.5	14.4	15.2	14.5	14.6	13.6	16.6	15.4	14.7 (-24.2%)
	EB-Net† [9]	24.8	27.6	25.6	28.3	25.9	28.9	29.4	29.3	30.5	32.0	32.8	26.1	28.6	26.3	27.9	29.2	28.6	30.8	31.8	27.2	28.6 (-22.1%)
	HiKER-SGG	**30.1**	**31.7**	**30.4**	**33.2**	**28.3**	**33.3**	**32.1**	**34.1**	**34.4**	**37.3**	**37.4**	**28.8**	**31.7**	**30.1**	**32.9**	**32.5**	**32.2**	**34.5**	**36.7**	**30.4**	**32.6 (-17.0%)**
mR@100: C/UC	GB-Net† [85]	40.1	41.9	40.1	43.8	37.8	42.9	42.7	45.1	47.1	50.8	51.7	37.8	42.8	42.9	46.6	42.5	46.1	41.2	49.6	45.9	44.0 (-20.6%)
	EB-Net† [9]	54.7	56.0	52.9	56.8	52.4	55.6	55.3	58.4	59.9	61.6	61.1	53.3	55.0	54.3	57.7	56.4	57.6	59.0	60.7	54.8	56.7 (-14.5%)
	HiKER-SGG	**59.3**	**60.3**	**58.6**	**62.3**	**55.6**	**61.9**	**59.8**	**63.4**	**64.0**	**66.9**	**67.4**	**56.4**	**60.1**	**58.4**	**62.3**	**59.8**	**62.1**	**63.7**	**66.3**	**58.9**	**61.4 (-11.3%)**
	GB-Net† [85]	14.8	15.1	14.6	16.6	15.1	15.6	15.6	16.9	17.1	19.1	19.0	14.4	16.0	16.0	16.8	16.1	16.1	15.0	18.1	17.0	16.3 (-22.0%)
	EB-Net† [9]	28.7	30.1	27.8	31.9	27.1	31.1	30.5	32.8	32.4	36.1	35.7	28.2	30.9	28.4	30.9	31.4	31.0	31.8	33.9	29.6	31.0 (-20.9%)
	HiKER-SGG	**32.7**	**33.8**	**32.6**	**36.0**	**30.4**	**35.7**	**34.7**	**36.3**	**36.7**	**39.9**	**39.7**	**31.1**	**34.2**	**32.7**	**35.4**	**34.9**	**35.4**	**37.1**	**39.2**	**32.6**	**35.1 (-14.8%)**

Contact: Ce Zhang (cezhang@cs.cmu.edu), Yaqi Xie (yaqixie@cmu.edu)