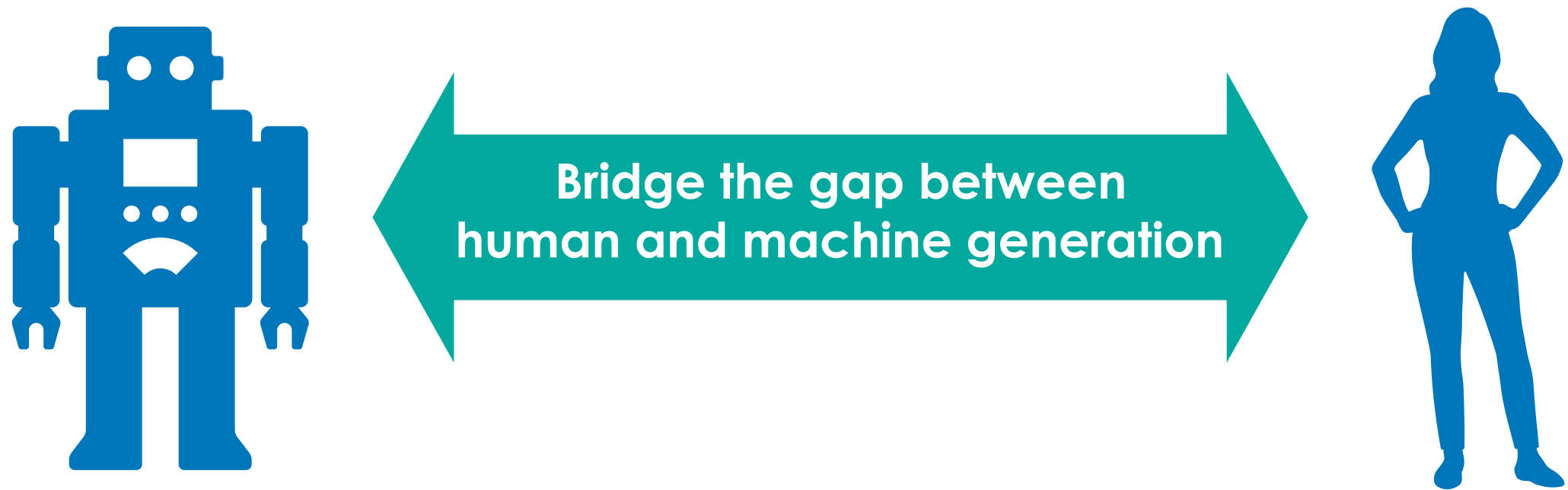


Entity Skeletons for Visual Storytelling

Ruo-Ping*, Khyathi Chandu*, Alan W Black



What makes a narrative effective?

Content - Relevance



We went to the beach.
My kids had a lot of fun there.
There were a lot of palm trees.
We stayed in a resort.



We went to the library.
I love reading books.
I borrowed a lot of them.

Content - Relevance



Entities

- Beach
- Kids
- Resort

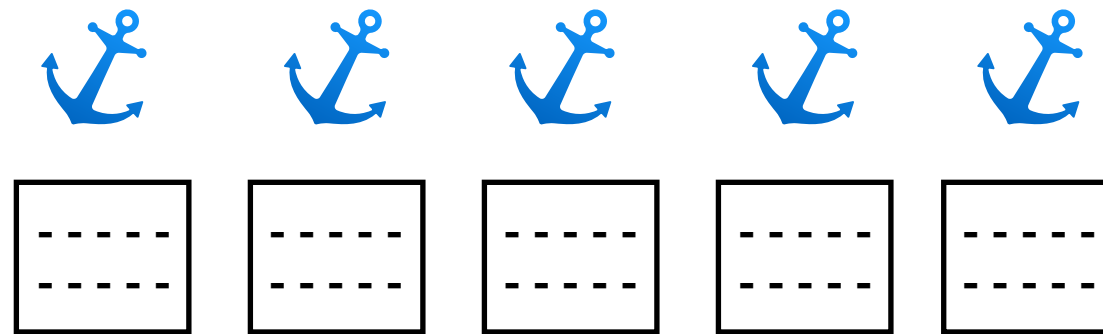


Entities

- Library
- Student
- Books

Anchoring Framework

Fine-grained Entity Skeleton



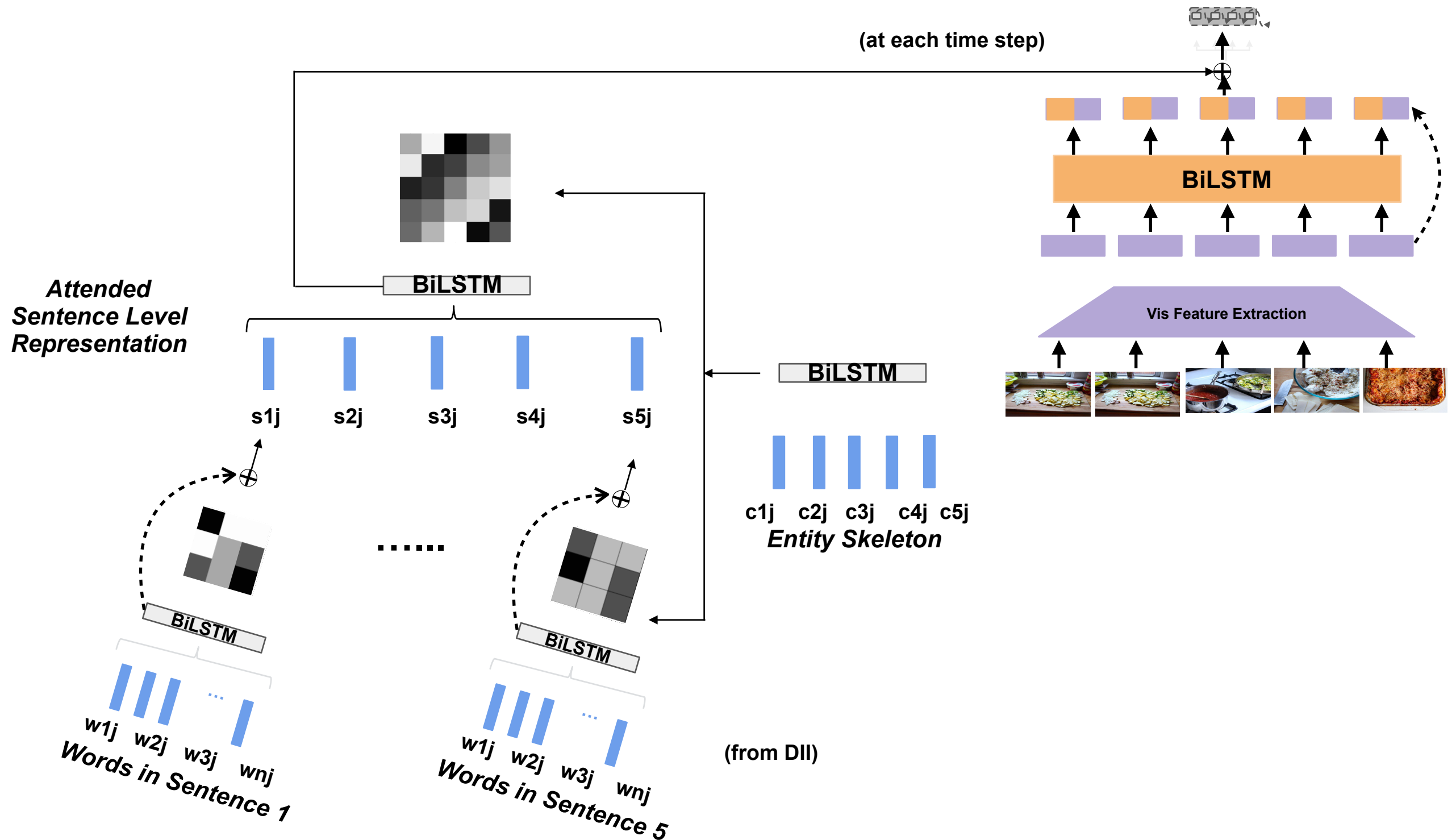
Input : I_i and $E_i = \{e_i^{(1)}, e_i^{(2)}, \dots, e_i^{(k)}\}$

Output : $N_i = \{s_i^{(1)}, s_i^{(2)}, \dots, s_i^{(k)}\}$

Provides full guidance to each individual unit of narrative text

Anchor Informed Generation

Hierarchical Glocal Model



Takeaways

- Improves Relevance component of visual storytelling
- Improves Controllability in generation
- Step towards interpretability with respect to intermediate representation

Thank You

Contact: kchandu@cs.cmu.edu