UAlberta at SemEval-2023 Task 1: Context Augmentation and Translation for Visual WSD

Mike Ogezi, Bradley Hauer, Talgat Omarov, Ning Shi, Greg Kondrak

{mikeogezi,bmhauer,omarov,nshi2,gkondrak}@ualberta.ca

Takeaways & Main Findings

- **Augmenting context** significantly improves semantic understanding

- English-language bias extends to the **vision-language** case

Task

- Given a **<u>focus word</u>** in **context**, and a set of **candidate images**, determine which **image** best depicts the meaning of the word.

- Example:
  - **Focus word:** bat   **Context:** baseball <u>bat</u>

Method: Context Augmentation & Translation
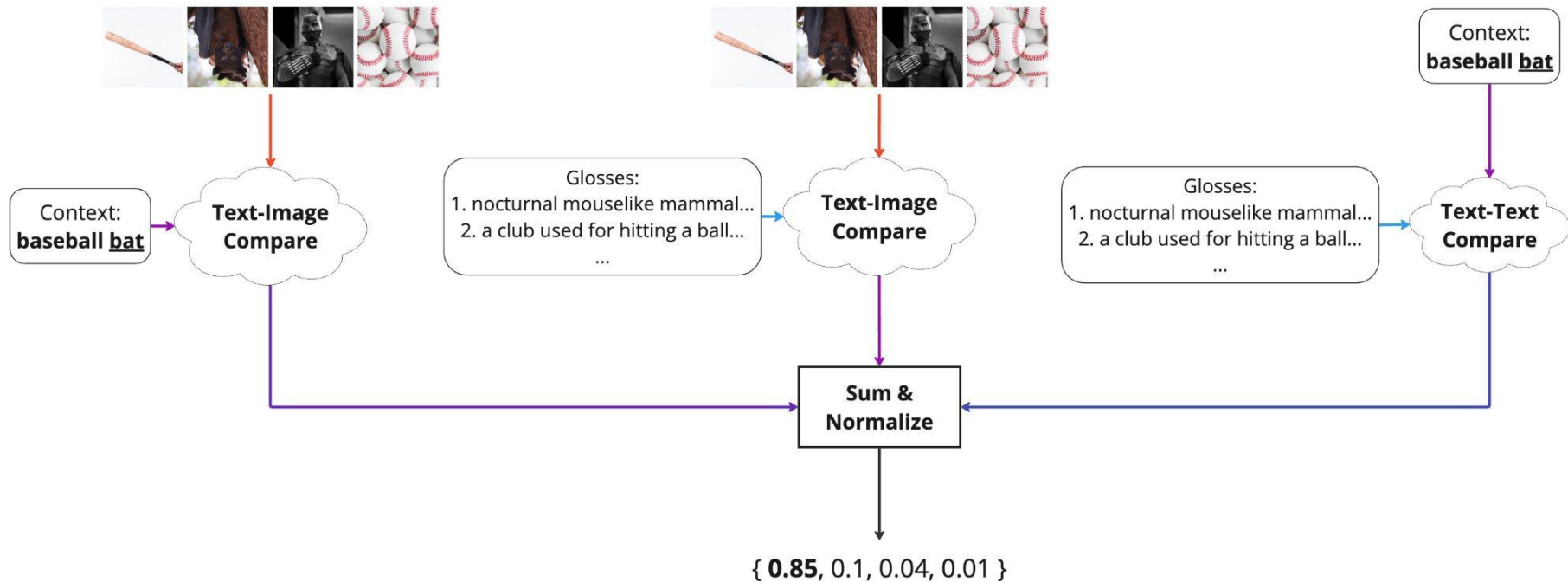
- **Context Augmentation:**

- **Translation:**

4

Method: Image Scoring

- **Intuition:** Pick **image** with strongest affinity to both the

  - **Context**, and

  - **Glosses** of the **senses** of the focus word

Method: Image Scoring

Experiments: Baselines vs Methods

- ## **Baseline:**

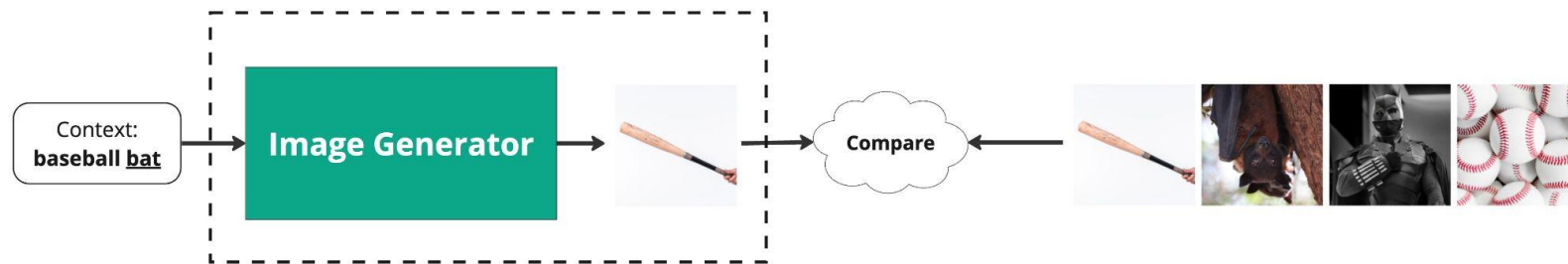   - Compare images to context with **multilingual** models

- ## **Language-Specific (LangSpec):**

    - Compare images to context with **language-specific** models
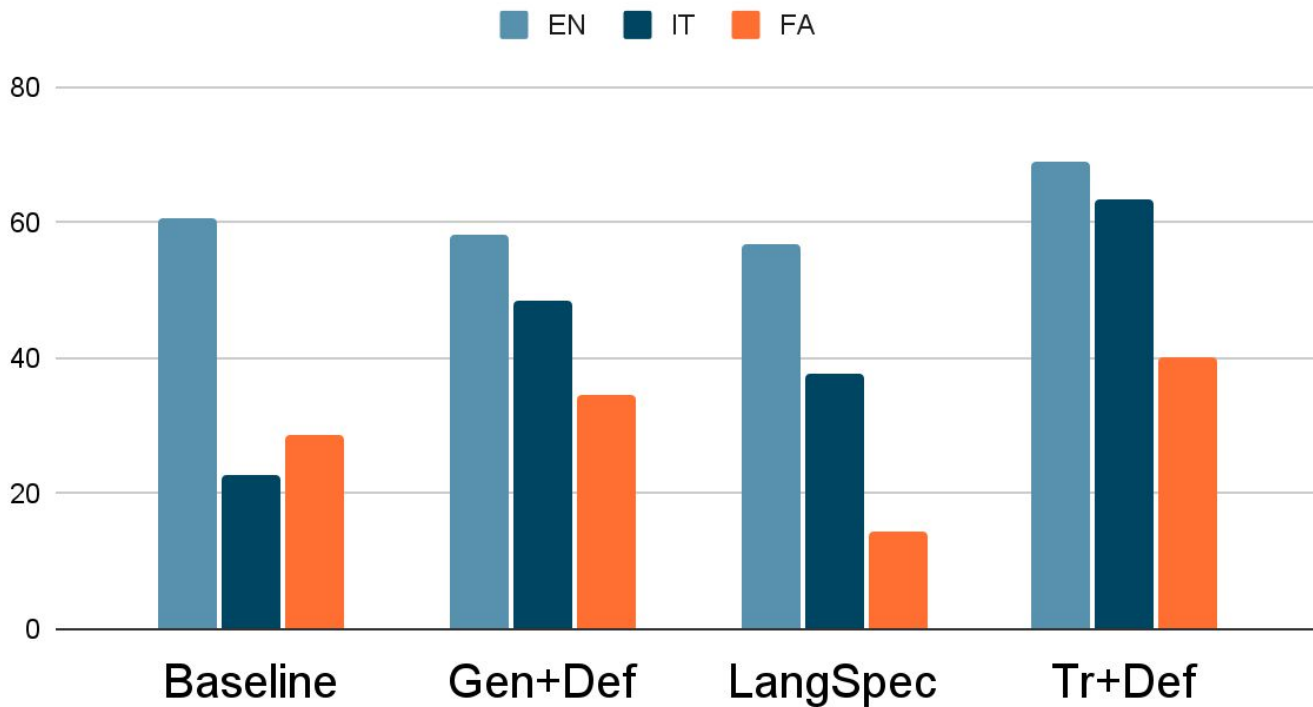
Experiments: Baselines vs Methods

- **Gen+Def: Image Generation & Context Augmentation:**
    - Compare candidate images to images **generated** from context

- **Tr+Def: Translation & Context Augmentation:**

    - Our primary method described earlier

Results: Accuracy on Test Set

Conclusions

- We find that **context augmentation** improves performance considerably in our task: **~20%** across all languages

- **English-only CLIP** on translated text, yields higher accuracy than **language-specific CLIP** on original text.