

Character-Aware Models Improve Visual Text Rendering

Rosanne Liu*, Dan Garrette*, Chitwan Saharia, William Chan, Adam Roberts,
Sharan Narang, Irina Blok, RJ Mical, Mohammad Norouzi, Noah Constant*

Google Research



Imagen



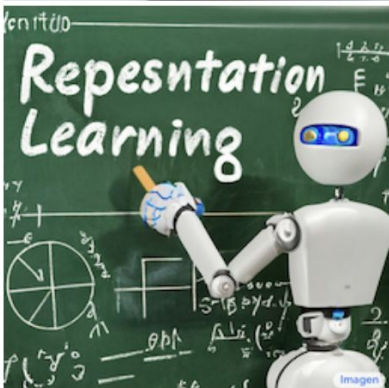
Imagen



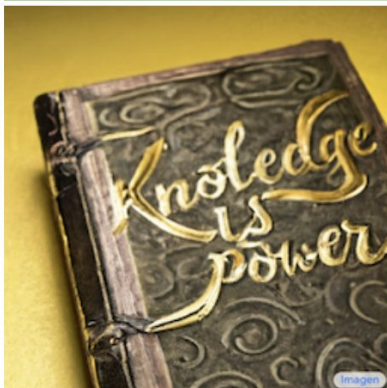
Imagen



Imagen



Imagen



Imagen



Imagen



Imagen

Text-to-Image Modeling

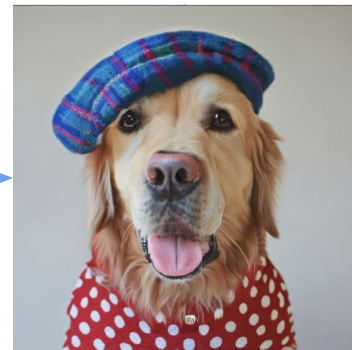
Imagen (Saharia et al., 2022)

A Golden Retriever
dog wearing a blue
checkered beret
and red dotted
turtleneck

input text

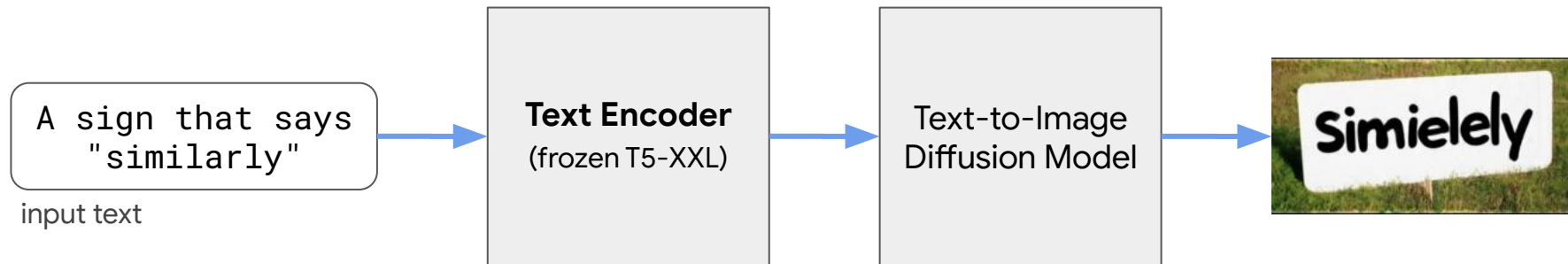
Text Encoder
(frozen T5-XXL)

Text-to-Image
Diffusion Model



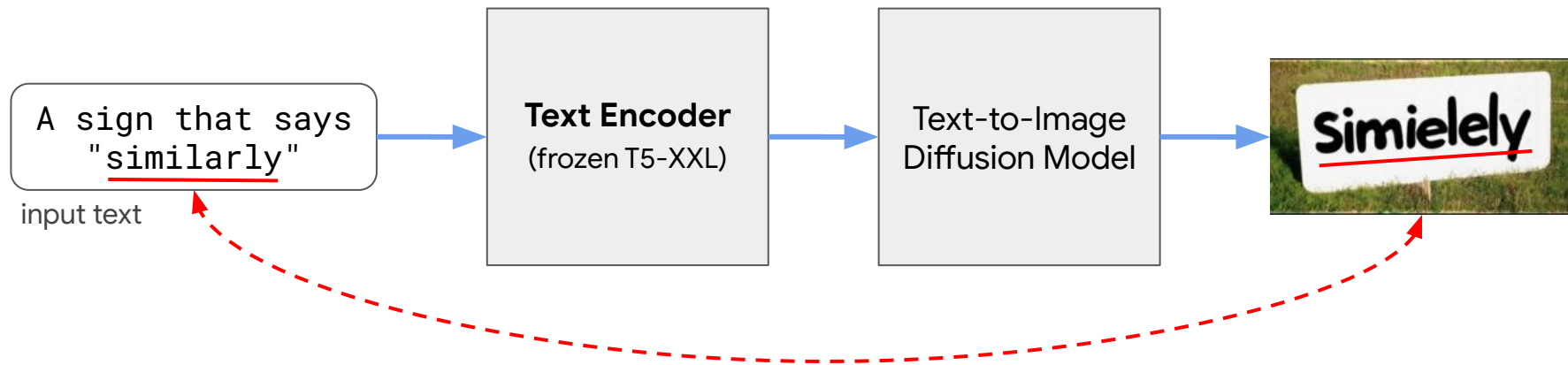
Text-to-Image Modeling

Imagen (Saharia et al., 2022)



Text-to-Image Modeling

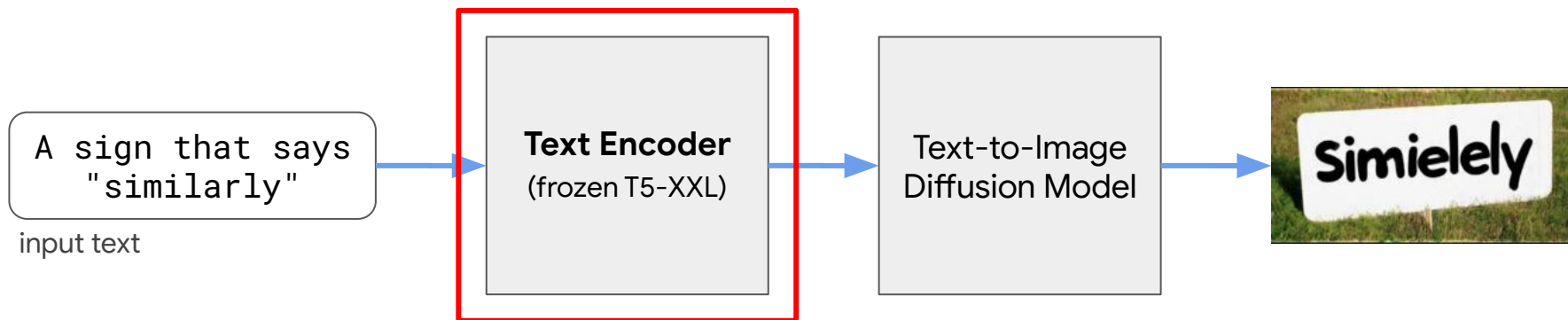
Imagen (Saharia et al., 2022)



Why doesn't the model just copy the spelling from the input?

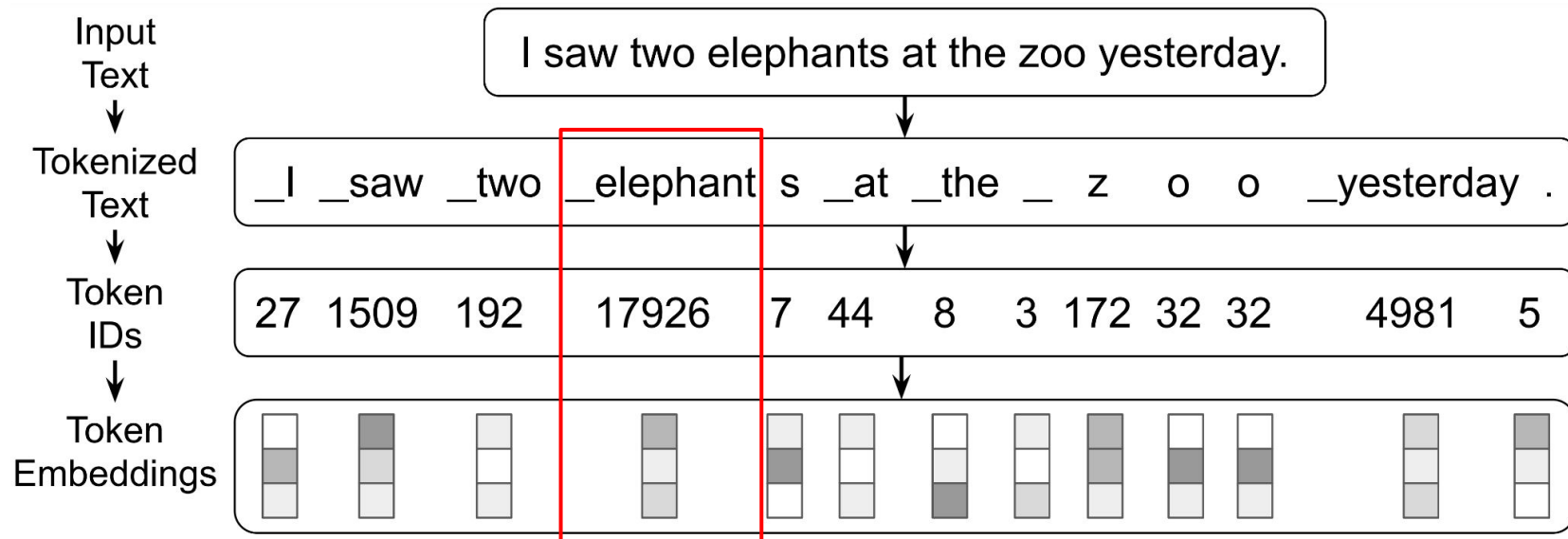
Text-to-Image Modeling

Imagen (Saharia et al., 2022)



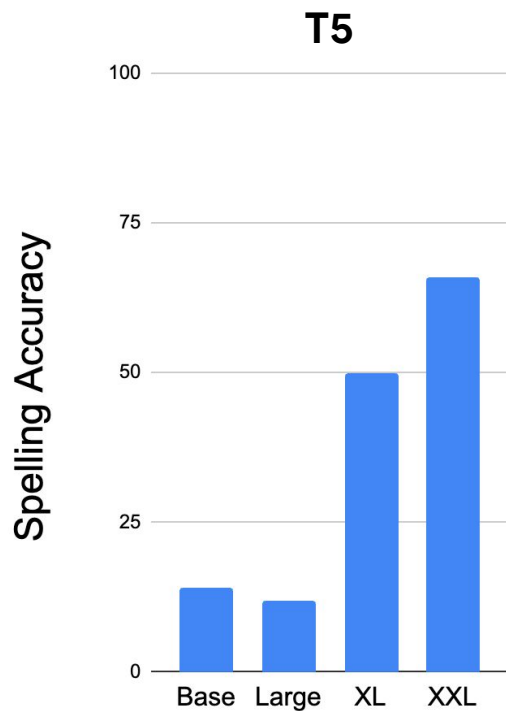
Tokenization Hides Spellings

SentencePiece tokenization in T5



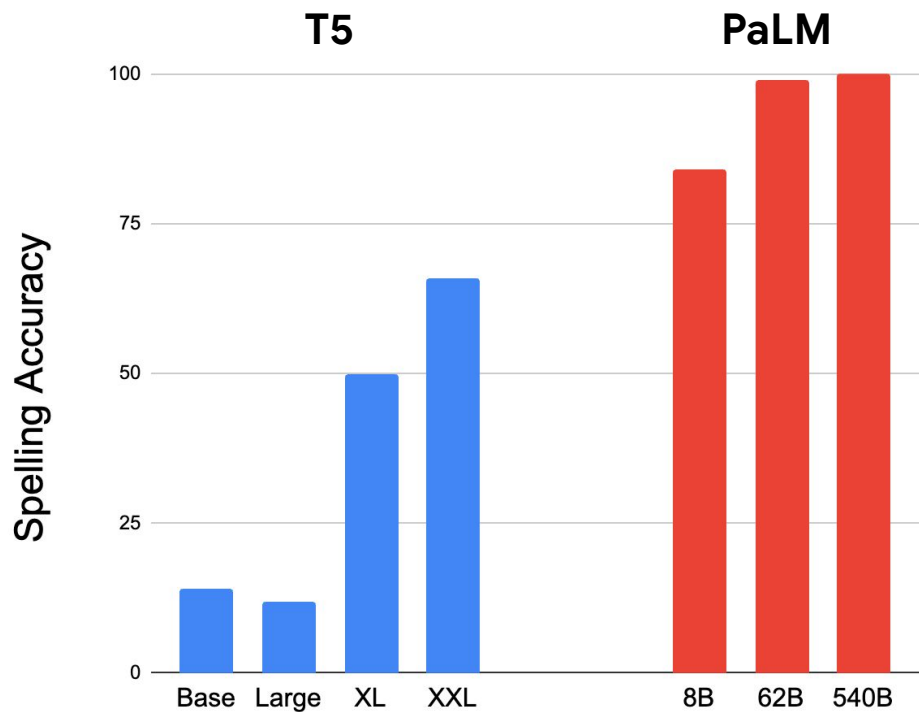
Does the text encoder know how to spell?

Subword-based encoder spelling ability depends on scale



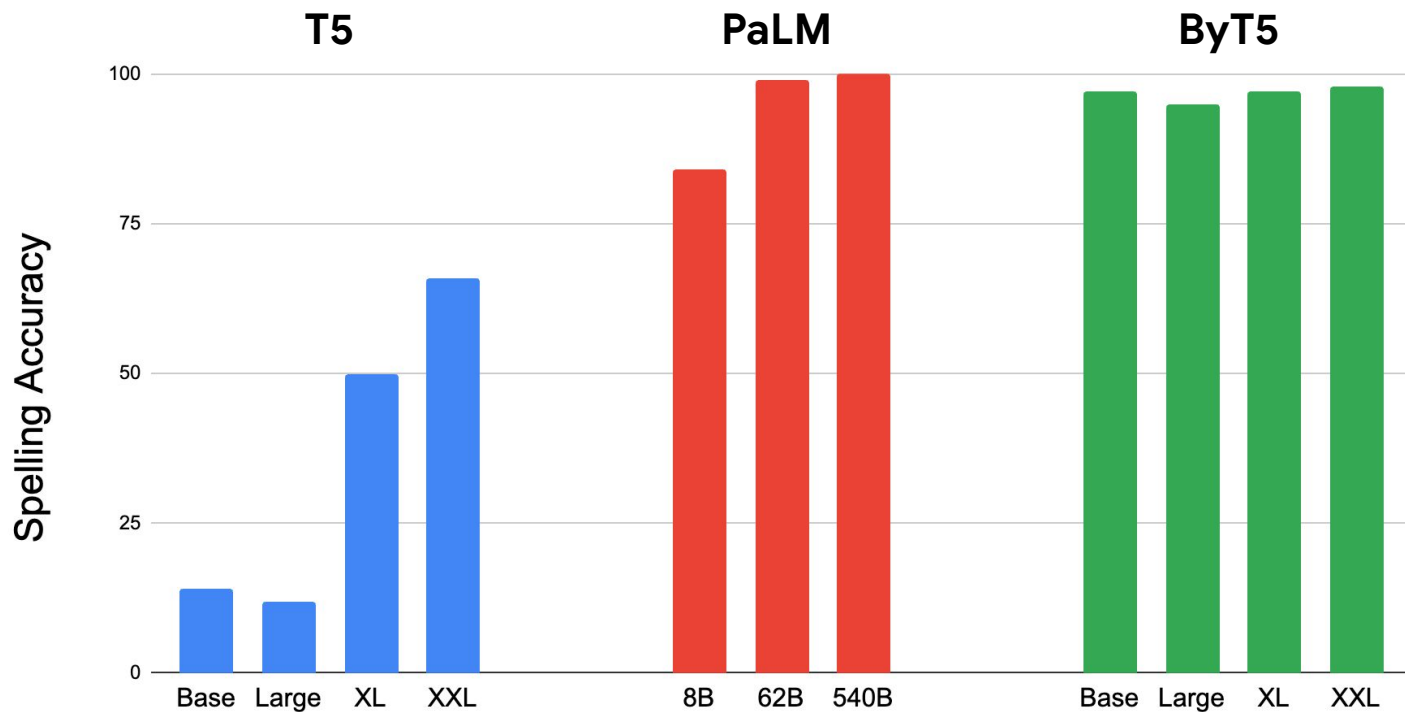
English, high-frequency words only.

Subword-based encoder spelling ability depends on scale



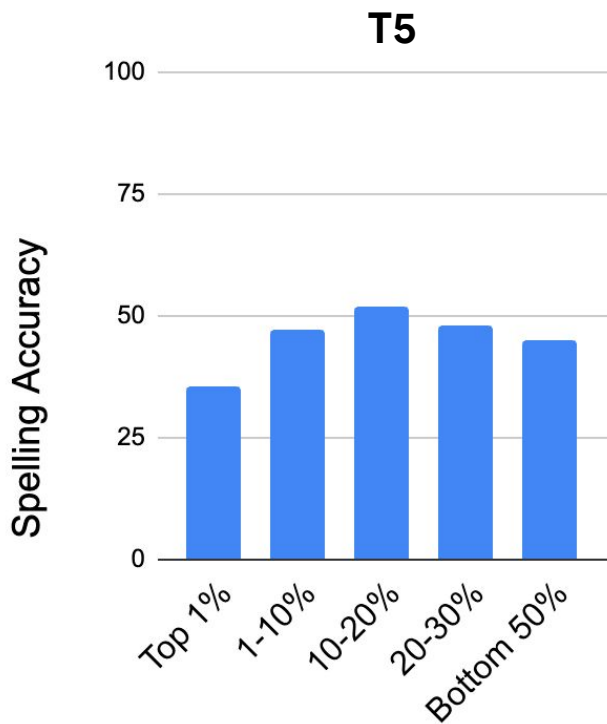
English, high-frequency words only.

Character-aware text encoders spell well at all scales



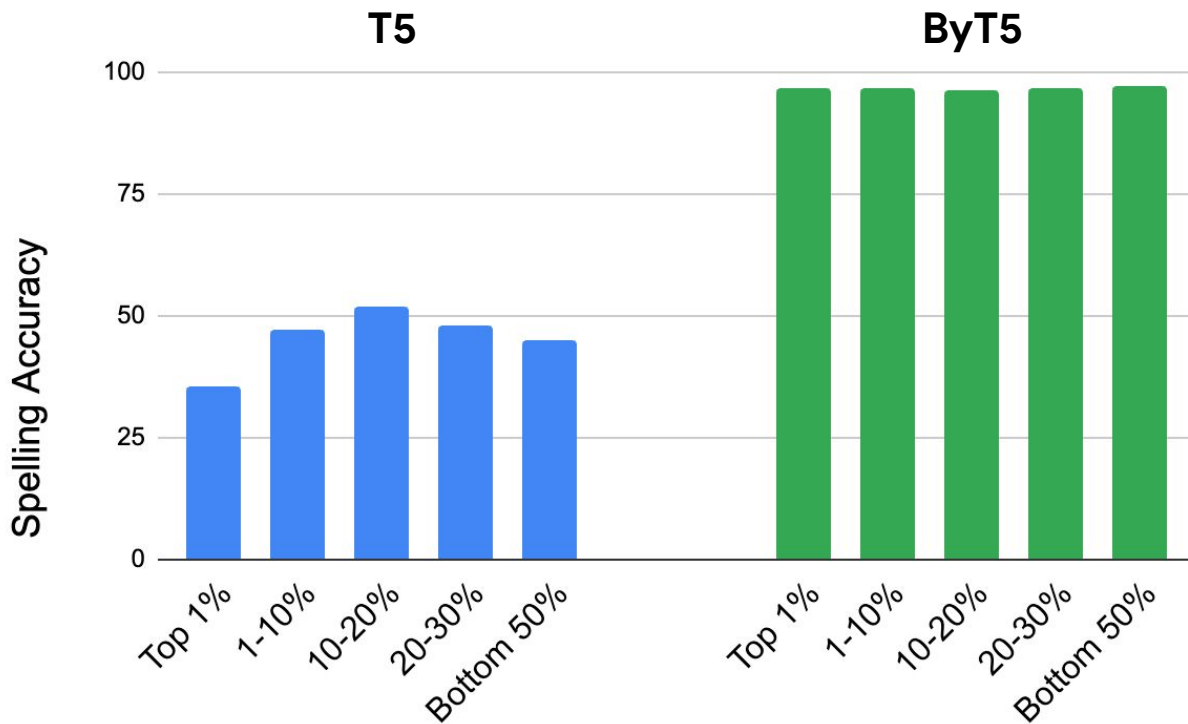
English, high-frequency words only.

Subword-based encoders are affected by word frequency



English words only. Results averaged across model sizes.

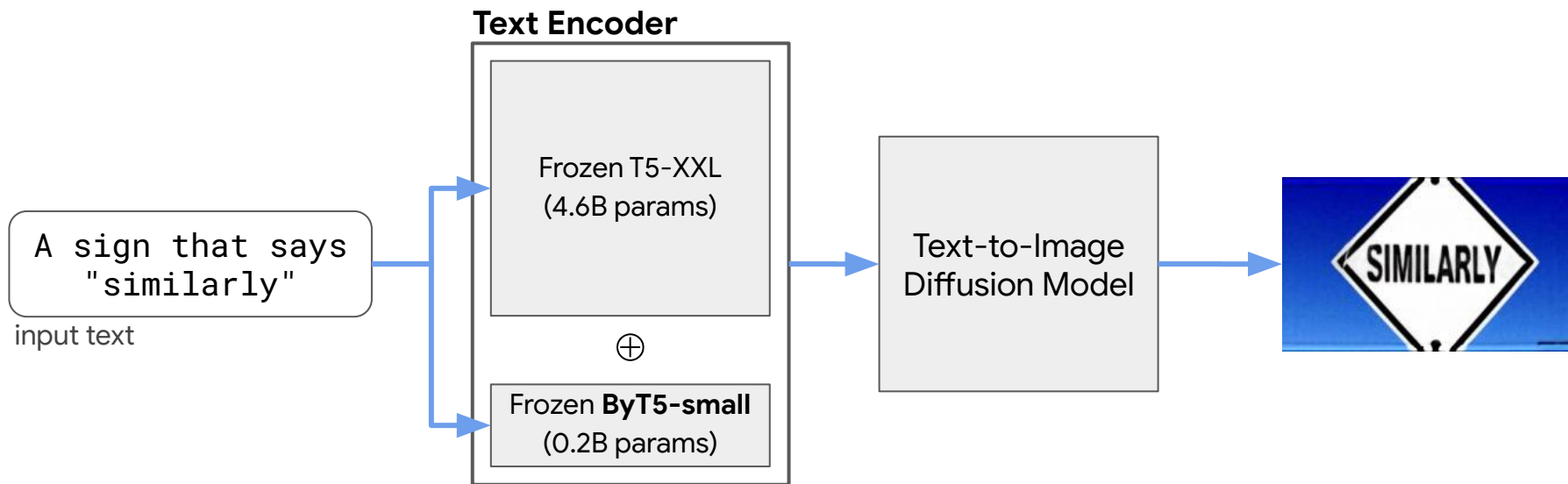
Character-aware encoders spell well across frequencies



English words only. Results averaged across model sizes.

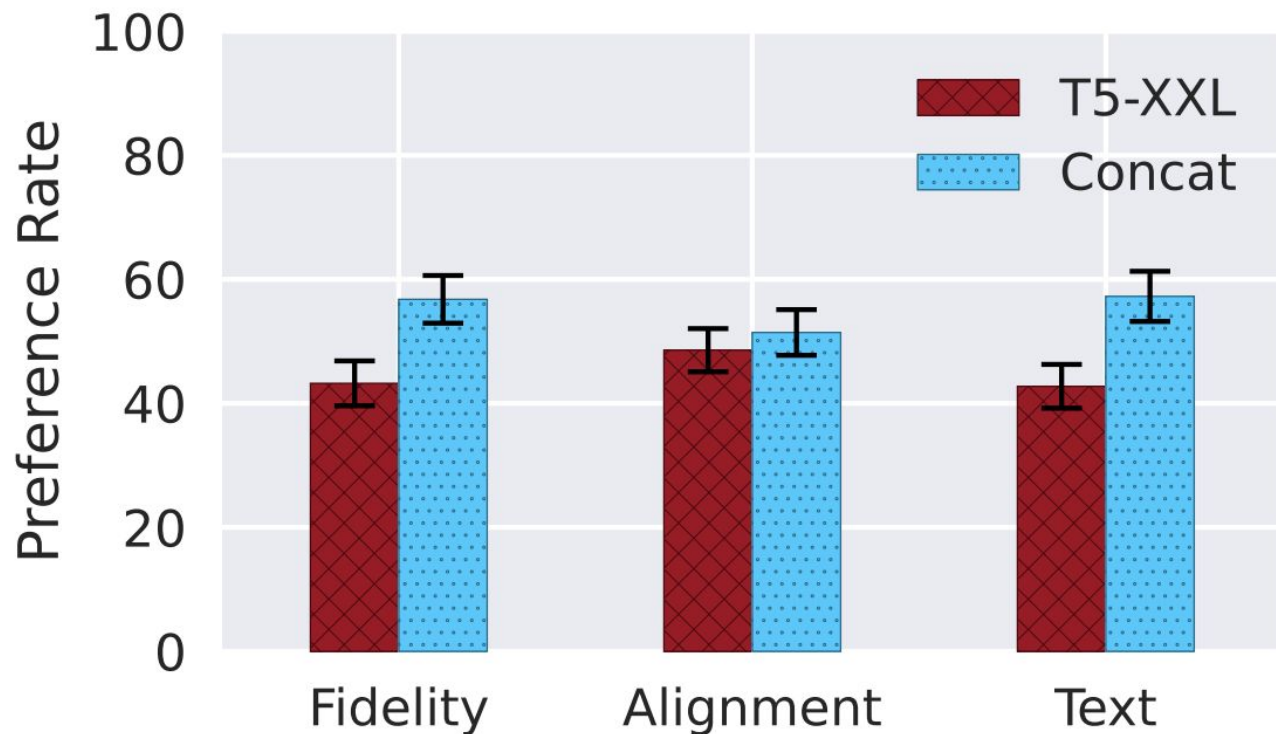
How can we improve text rendering?

Adding Character Information

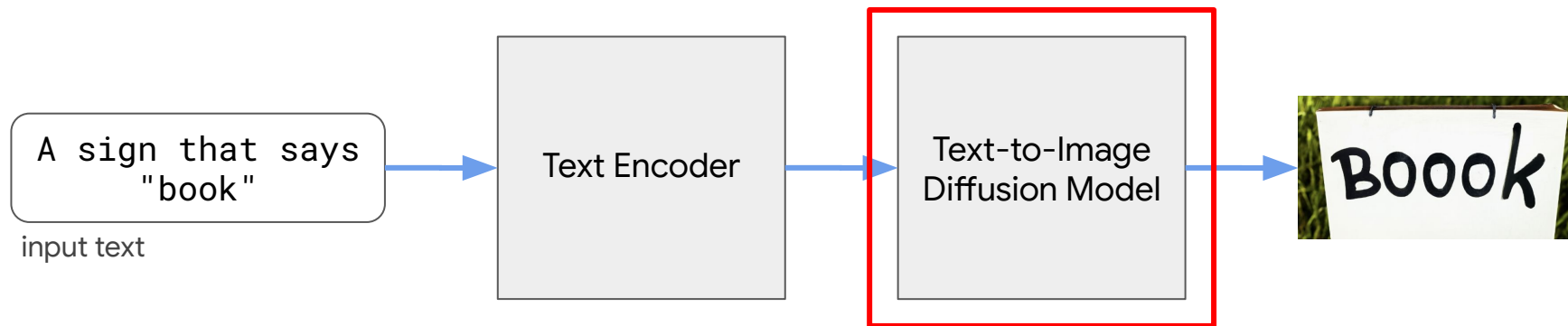


Concatenate subword-level and character-level text encodings

Character-awareness improves image generation metrics



Text-to-Image Modeling



Errors during image generation

Excess repetitions



Merged glyphs



Misshapen glyphs



No text



Takeaways

- WikiSpell – Benchmark for text-only models.
- DrawText – Benchmark for text-to-image models.
- Efficient strategy for improving model spelling ability.

Takeaways

- WikiSpell – Benchmark for text-only models.
- DrawText – Benchmark for text-to-image models.
- Efficient strategy for improving model spelling ability.

