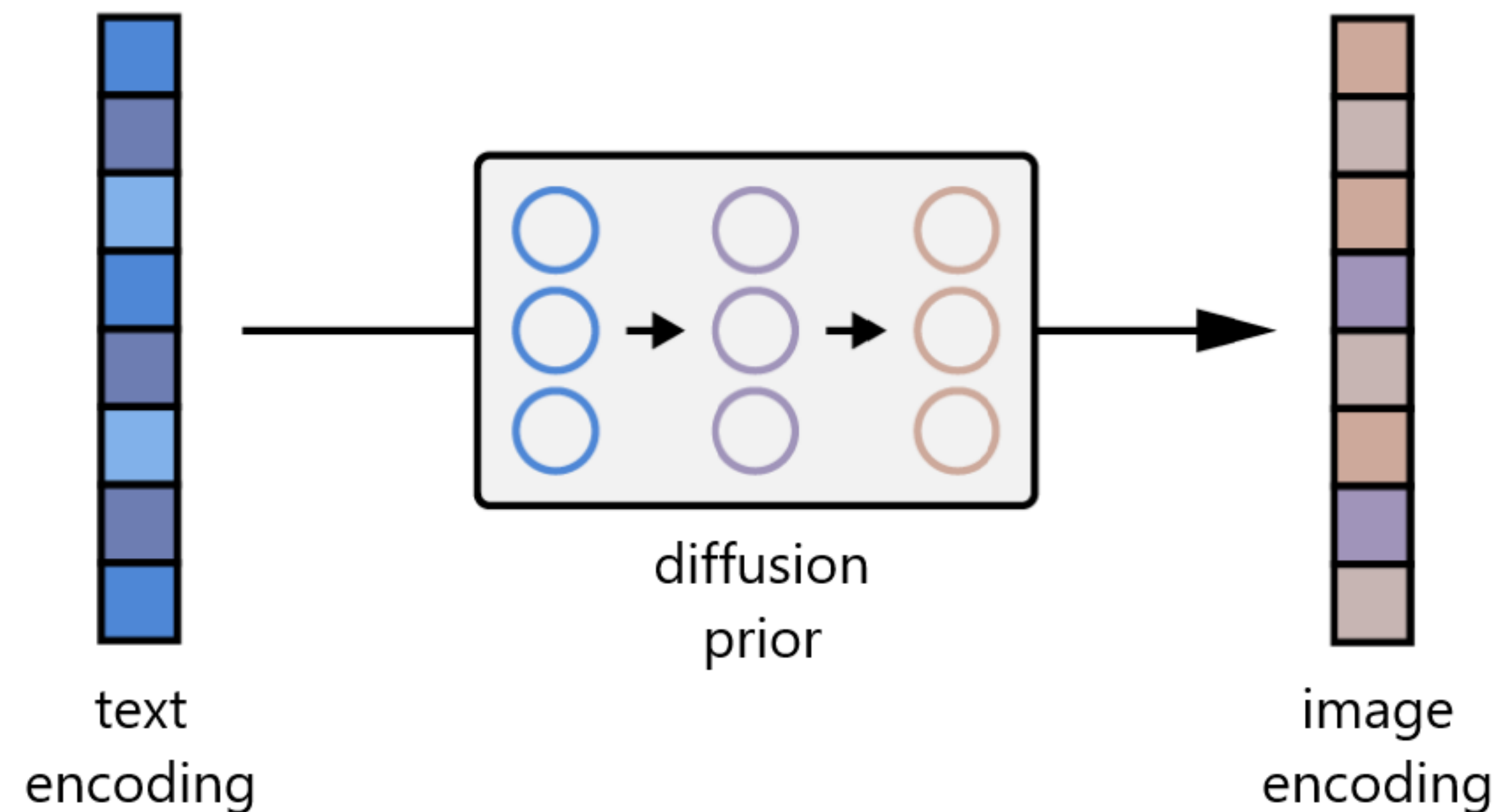


MAPPING FROM TEXT TO IMAGE

- Mapping from Textual Semantics to Corresponding Visual Semantics
- In addition to our image encoder, CLIP also learns a text encoder. DALL-E 2 uses another model, which the authors call the prior, in order to map from the text encodings of image captions to the image encodings of their corresponding images



BUILDING IT ALL

