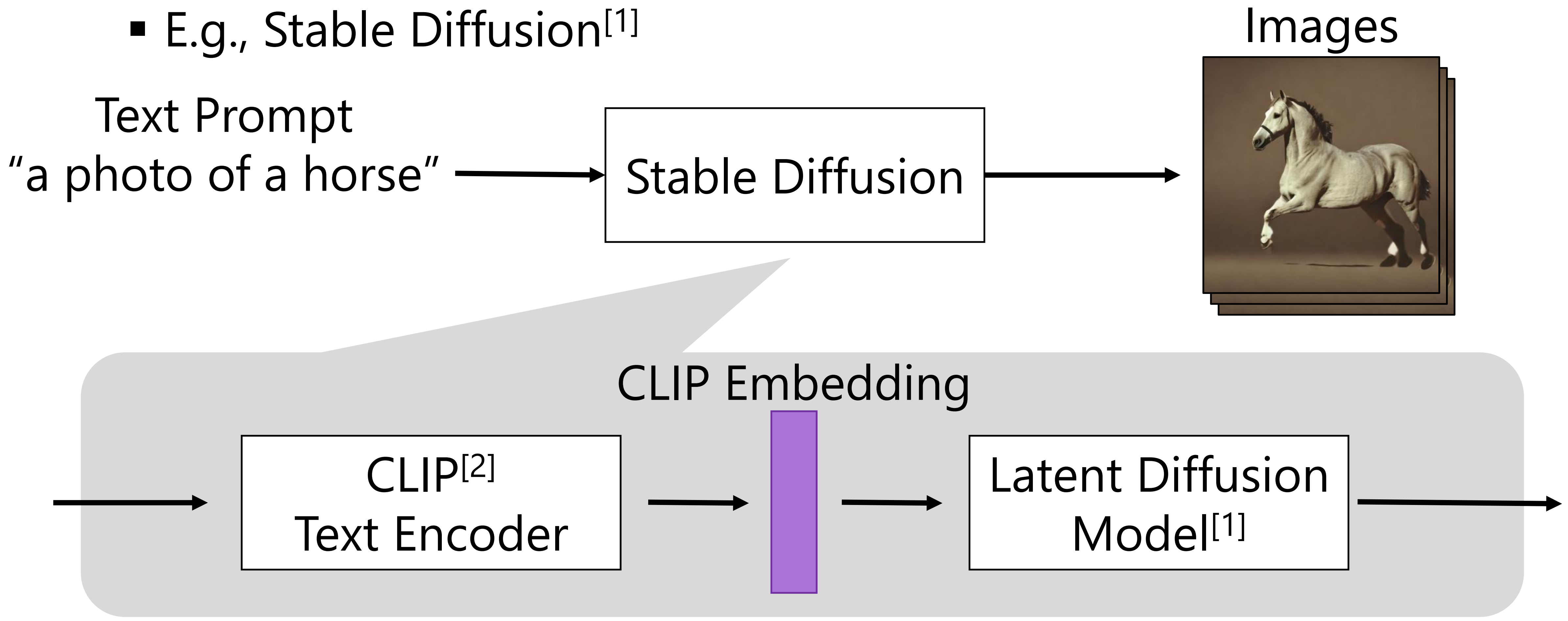
32nd ACM Multimedia Conference (ACM MM’24), October 31, 2024
Paper 2993
OS18-6

Investigating Conceptual Blending of a Diffusion Model for Improving Nonword-to-Image Generation

Chihaya Matsuhira, Marc A. Kastner, Takahiro Komamizu, Takatsugu Hirayama, Ichiro Ide
Nagoya University, Japan

Text-to-Image Diffusion Models

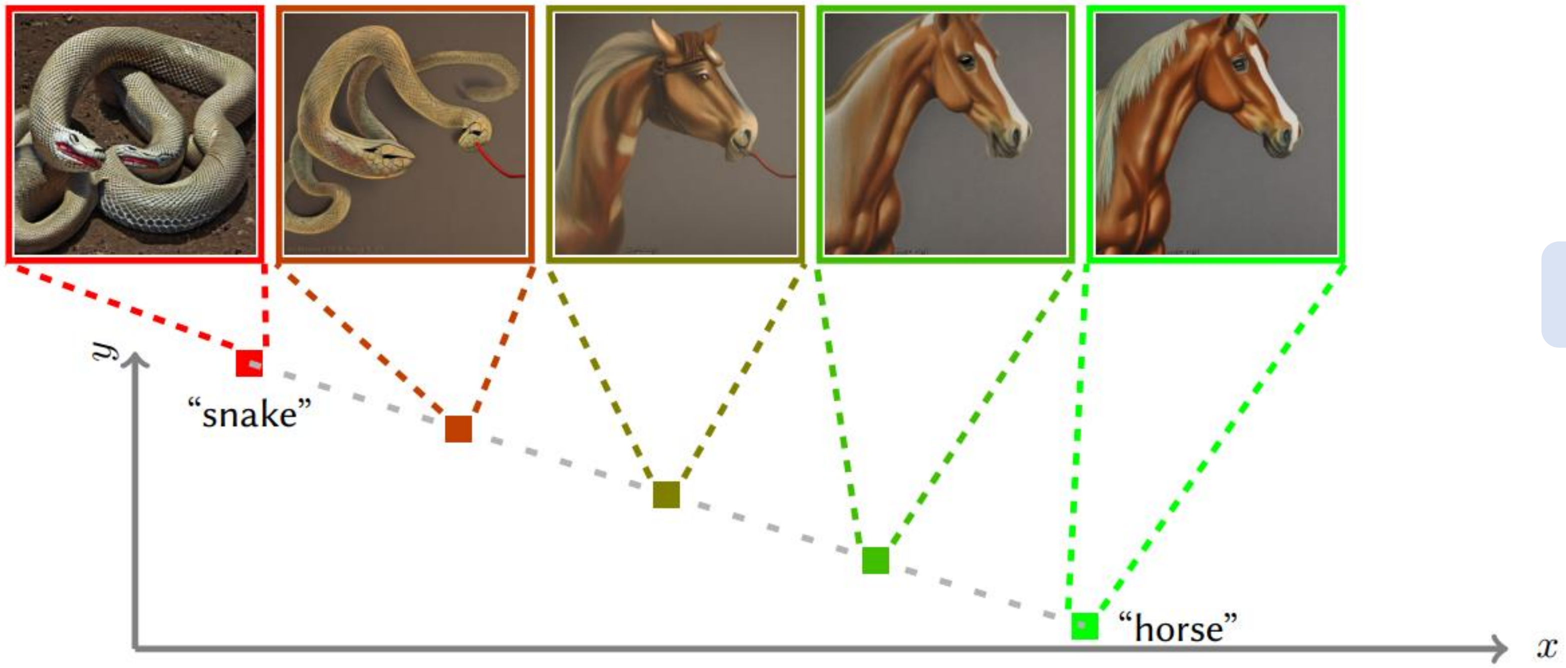
- Enable image generation from a text prompt
  - E.g., Stable Diffusion[1]



[1] Rombach et al., "High-resolution image synthesis with latent diffusion models", CVPR 2022.
[2] Radford et al., "Learning transferable visual models from natural language supervision", ICML 2021.

Conceptual Blending in Diffusion Models

- Melzi et al.[3] assessed this effect **qualitatively**
  - **Interpolated embedding** between two concepts induces images depicting blended concepts



CLIP Embedding Space (Illustrated by Melzi et al.[3])

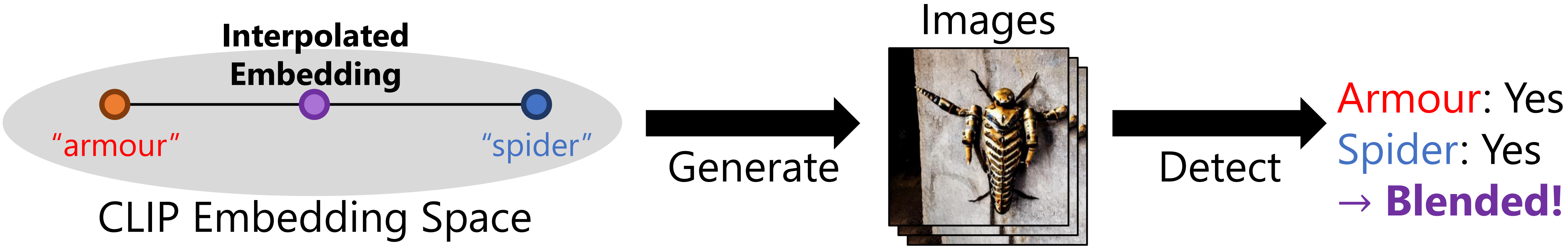
Problem

No quantitative results

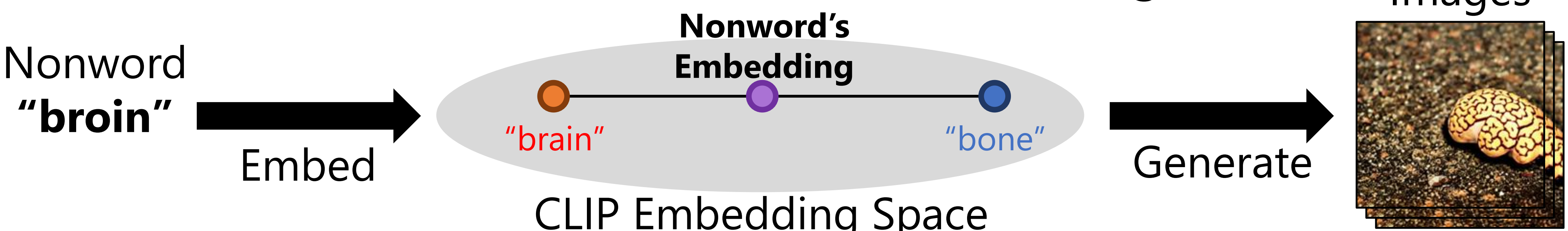
[3] Melzi et al., “Does Stable Diffusion dream of electric sheep?”, Image Schema Day 2023.

Two Contributions of This Paper

1. Evaluate **conceptual blending** quantitatively
   - Investigate when and how often two concepts blend



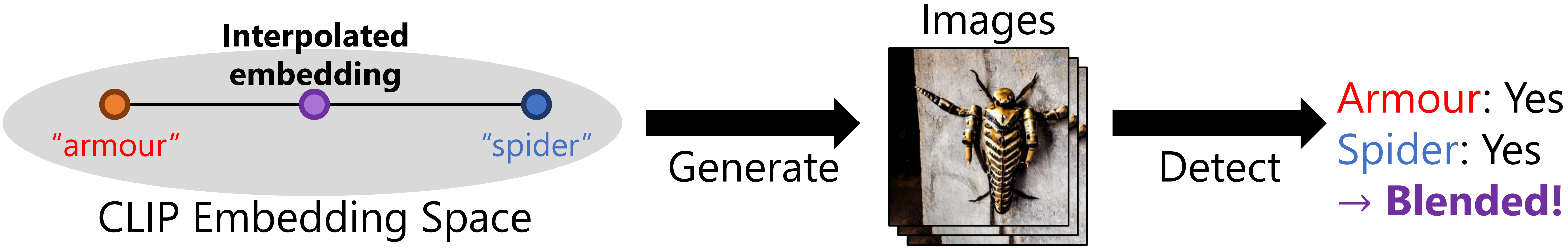
2. Exploit conceptual blending to generate intuitive images for **non-existing words (nonwords)**
   - Associate each nonword with similar-sounding words



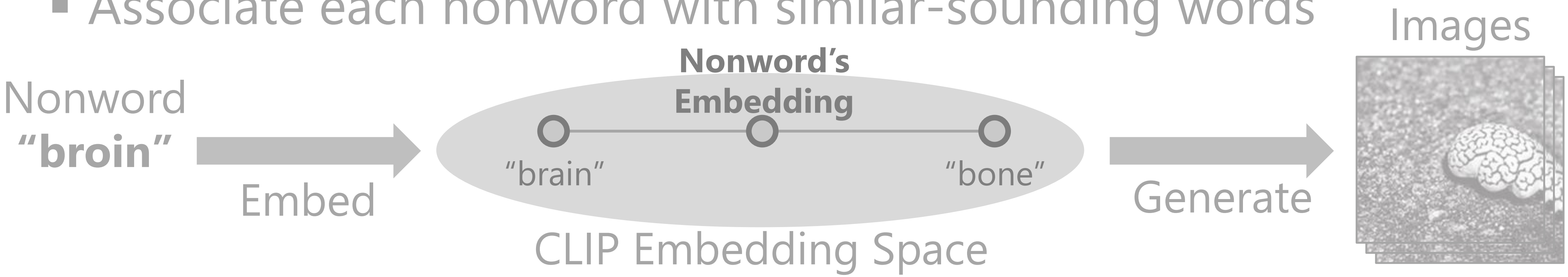
Focus of This Oral Presentation

Rest will be explained in our poster!

1. Evaluate **conceptual blending** quantitatively
   - Investigate when and how often two concepts blend

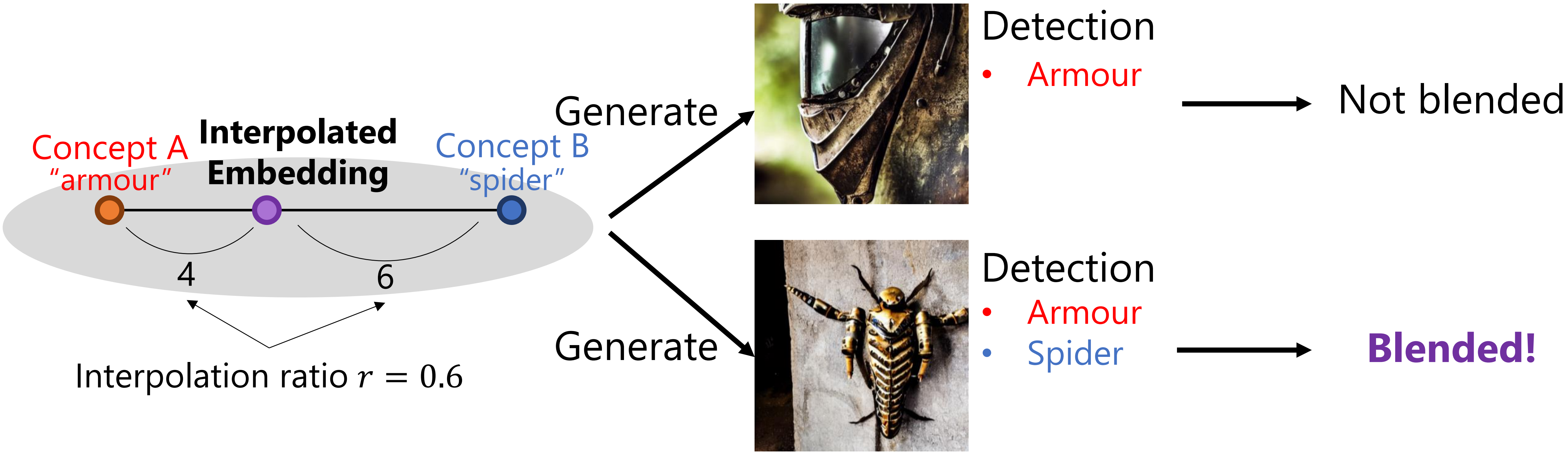


2. Exploit conceptual blending to generate intuitive images for **non-existing words (nonwords)**
   - Associate each nonword with similar-sounding words



Evaluating Conceptual Blending Quantitatively

- Purpose: Detect conceptual blending in generated images
- Approach: Construct a concept detector
  - Apply it to images generated from an **interpolated embedding**
    - ➢ Blending is the case where two concepts are detected simultaneously



Constructing Concept Detector using CLIP

- Use CLIP[2] as an open-vocabulary concept detector

Generated image 1

CLIP[2]

CLIP Score

Armour

Spider

Thresholding

Detected Concept

Armour

Not blended

Concepts to Detect

Armour / **Spider**

Generated image 2

CLIP[2]

CLIP Score

Armour

Spider

Thresholding

Detected Concept

Armour / **Spider**

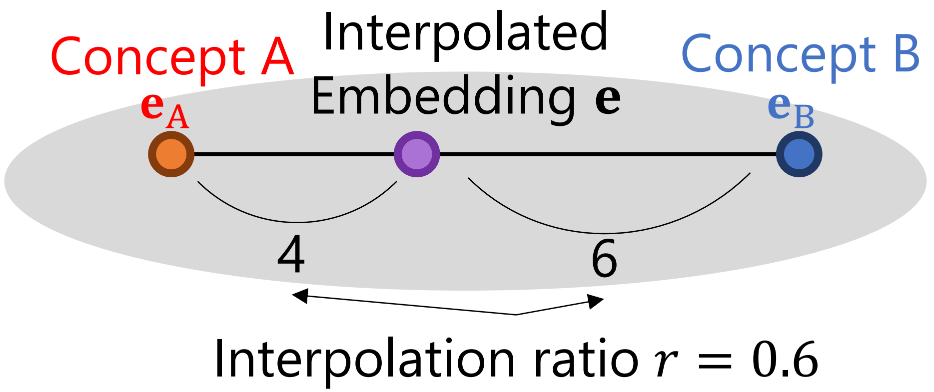
Blended!

[2] Radford et al., "Learning transferable visual models from natural language supervision", ICML 2021.

Data: Preparing Seed Concepts

- Seed Concepts: English nouns with high imageability
  - Imageability: Measure of how easily a word can be imagined
  - Randomly taken from MRC Psycholinguistic Database[4]

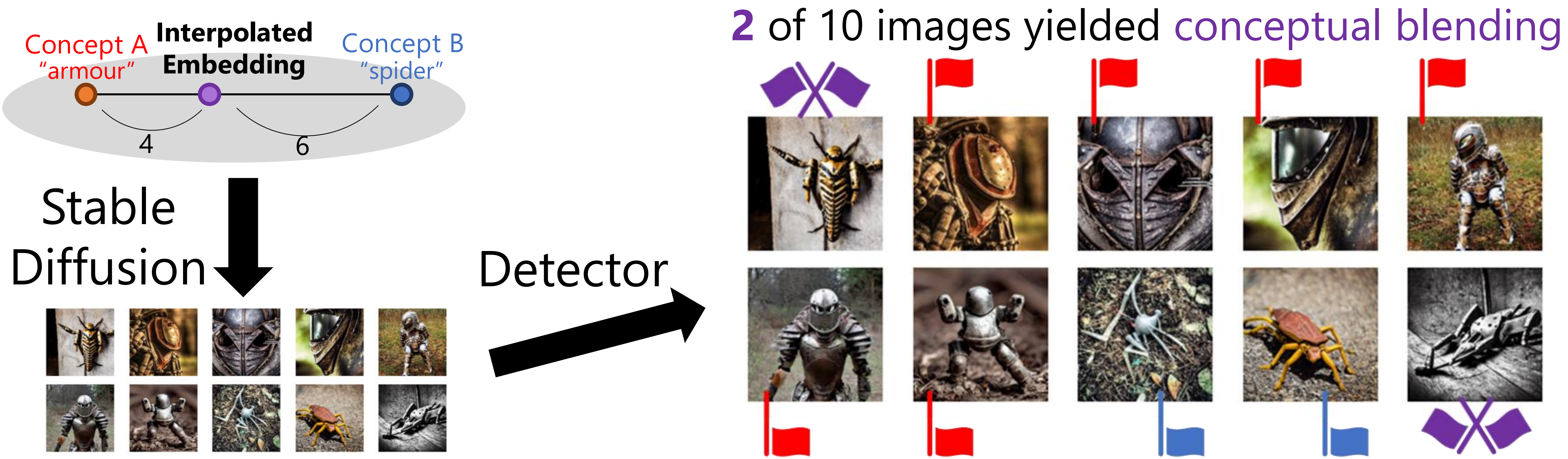
- Interpolated embeddings between concepts A and B
  - Formula: $\mathbf{e} = r\mathbf{e}_{\mathrm{A}} + (1-r)\mathbf{e}_{\mathrm{B}}$
    - r: Interpolation ratio
      - Randomly sample from {0.1, 0.2, ..., 0.9}
  - Created by randomly sampling two seed concepts A and B



[4] Coltheart, "The MRC psycholinguistic database", Q. J. Exp. Psychol. Section A, vol.33, no.4, 1981.

Data: Interpolated Embeddings and Images

- Create 1,000 interpolated embeddings
  - By randomly sampling two seed concepts 1,000 times
- Generate 10 images for each interpolated embedding
- Detect blending with the concept detector



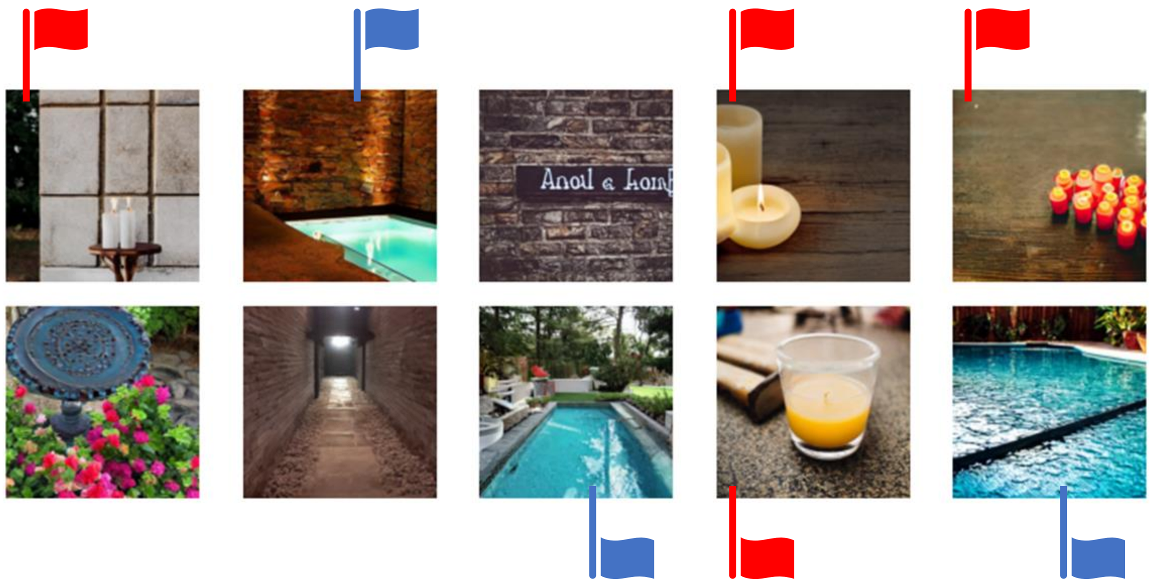
Metric: **Blended Concept Depiction (BCD)**

- Probability of detecting both concepts A and B in 2 or more images (out of 10 images)
  - Probability: Frequency divided by the total number of image sets

A: “armour”, B: “spider”

Counted as **BCD**

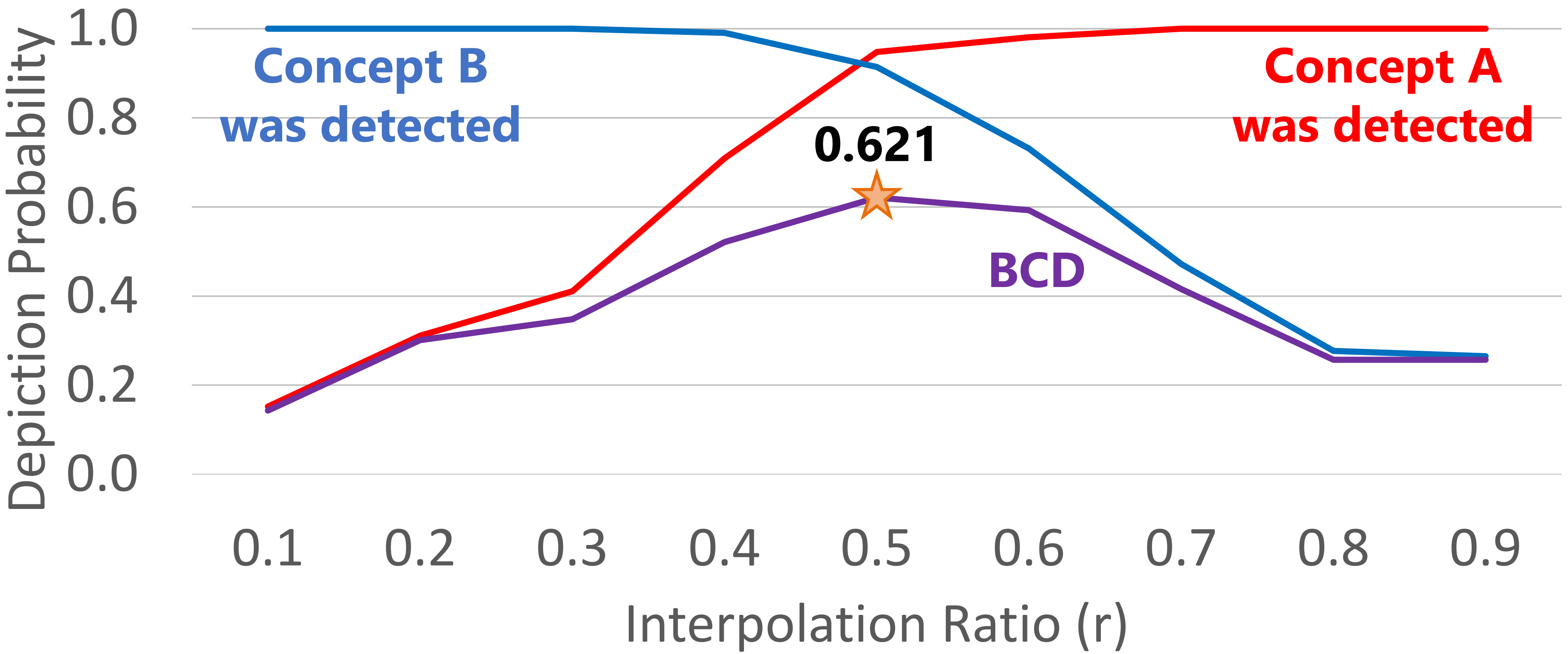
A: “candle”, B: “pool”



Not counted as **BCD**

Maximum 62.1% Image Sets Yield Blending

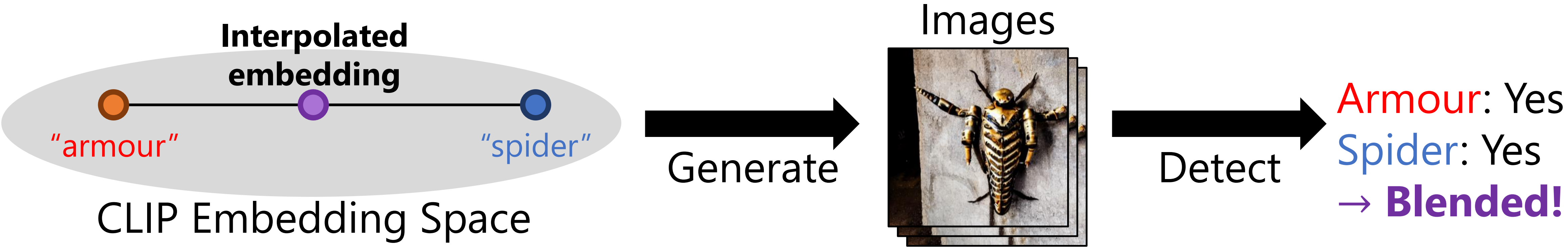
- More than **60%** image generation results depict blended concepts
- BCD probability increases as the interpolation ratio approximates **0.5**



Two Contributions of This Paper

Our Poster:
P047

1. Evaluate **conceptual blending** quantitatively
   - **Midpoint** maximizes the occurrence of conceptual blending



2. Exploit conceptual blending to generate intuitive images for **non-existing words (nonwords)**
   - Associate each nonword with similar-sounding words