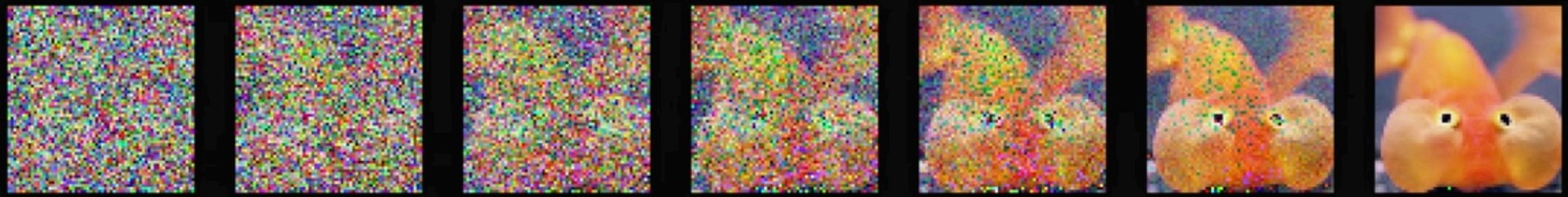
RELIC: Federated Conditional Textual Inversion with Prototype Alignment

Sijia Chen, Ningxin Su, Baochun Li

University of Toronto

“Blue sofa in a white room with a cactus to its right and a coffee table in front”



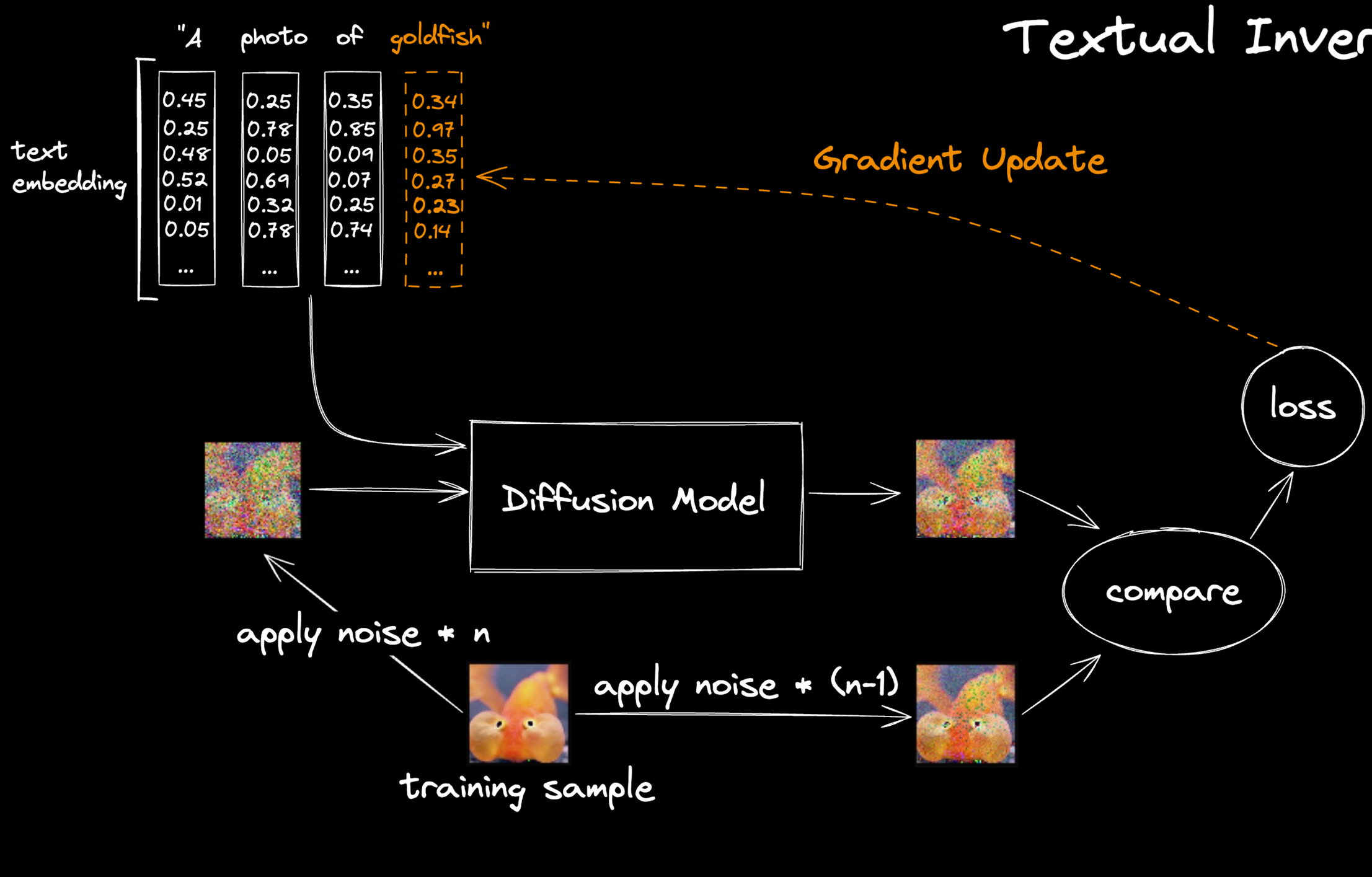
word2vec

a	-0.32, 0.64, -0.49 ...
aardvark	0.62, -0.65, -0.47 ...
abaca	0.07, -1.00, 0.56 ...
abalone	-0.25, 0.66, 0.35 ...
abandon	0.61, -0.67, 0.89 ...
⋮	
zounds	0.62, -0.65, -0.47 ...
zucchini	0.07, -1.00, 0.56 ...
zugzwang	0.13, 0.98, -0.71 ...
zwanziger	-0.09, 0.44, -0.73 ...
zwitter	-0.17, 0.43, 0.03 ...

Relationships within the same list: words that are more likely to appear **in similar contexts** will be more similar

Relationships within the same list: words that are more likely to appear **in similar contexts** will be more similar

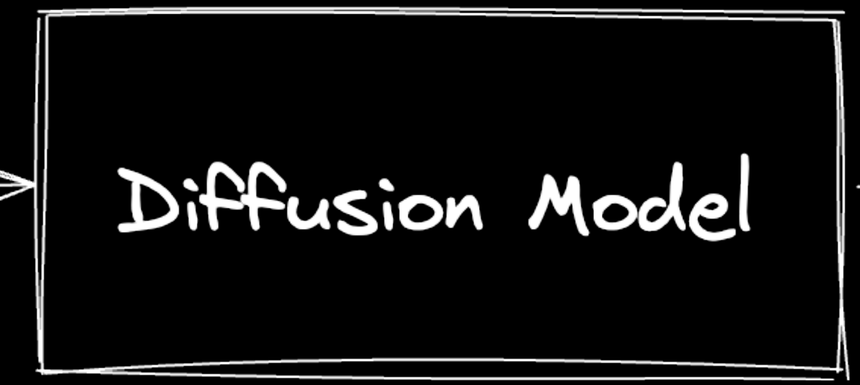
Textual Inversion



Relationships within the same list: words that are more likely to appear **in similar contexts** will be more similar

Federated textual inversion: aggregate pseudo-word embedding vector

"A photo of goldfish"



However, there is a privacy leakage problem.
An attacker can directly generate similar images to this client based on its pseudo-word


goldfish
0.34
0.97
0.35
0.27
0.23
0.14
...
Diffusion Model

The averaged pseudo-word embedding may lose learned features

Source images



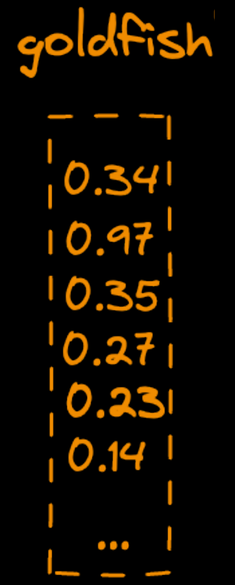
Prompt an oil painting of $S^\star$ a $S^\star$ themed lunchbox

TI-Central




FedTI



$$u' = W_1 \cdot F + W_2 \cdot u$$

Aggregated and updated u

K.Zhou,J.Yang,C.C.Loy,andZ.Liu,“Conditional Prompt Learningfor Vision-Language Models,” in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 16 816–16 825.

Improving performance further — prototype alignment

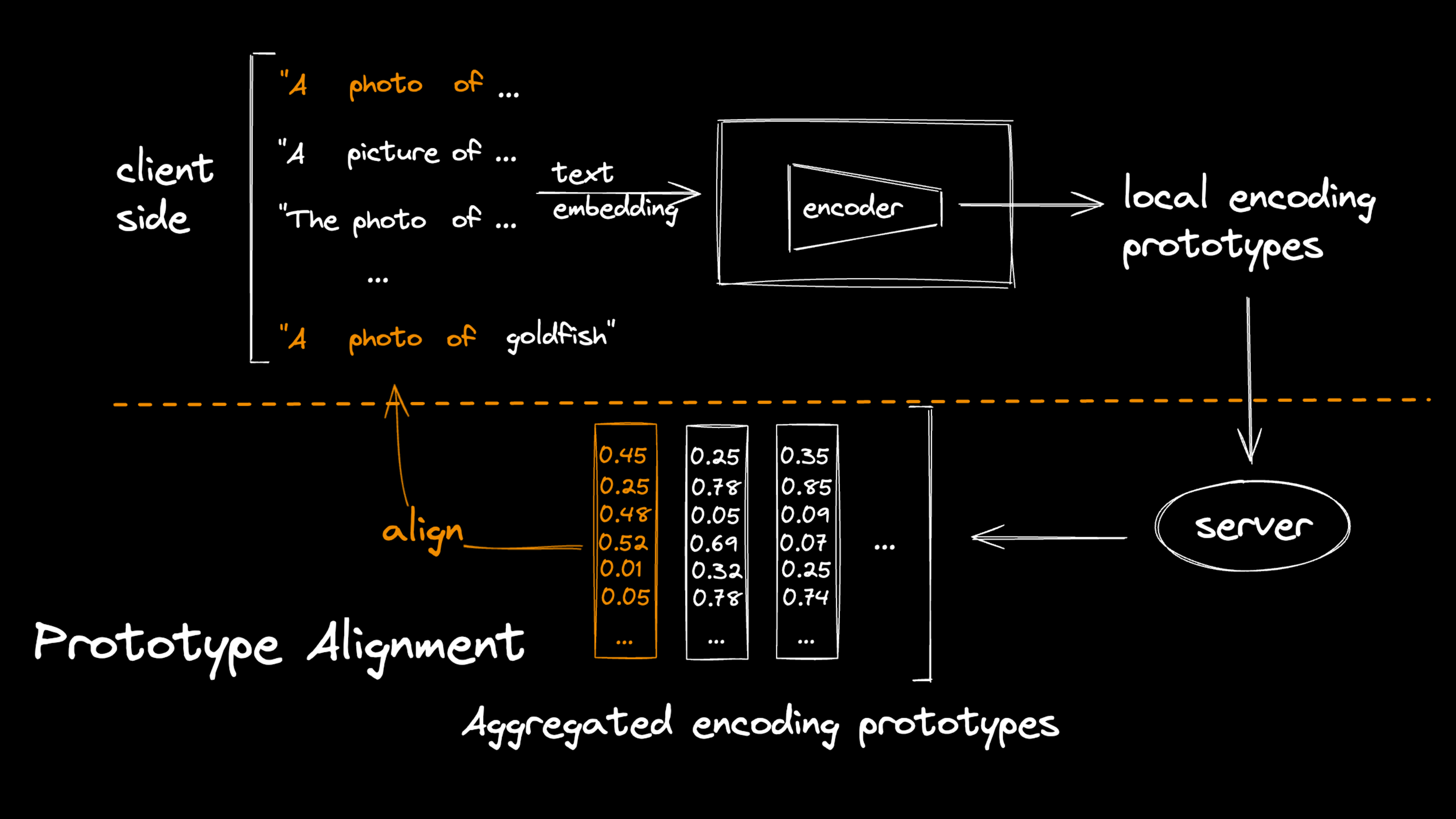
Source images



Prompt an oil painting of $S^\star$ a $S^\star$ themed lunchbox

FedCPW





client side
"A photo of ...
"A picture of ...
"The photo of ...
...
"A photo of goldfish"
text embedding
encoder
local encoding prototypes
server
align
0.45 0.25 0.48 0.52 0.01 0.05 ...
0.25 0.78 0.05 0.69 0.32 0.78 ...
0.35 0.85 0.09 0.07 0.25 0.74 ...
...
Prototype Alignment
Aggregated encoding prototypes

Source images

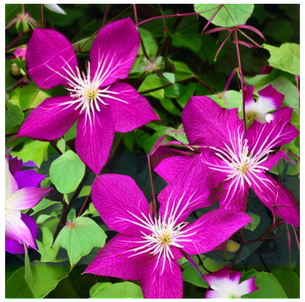


Prompt an oil painting of $S^\star$ a $S^\star$ themed lunchbox

FedCPW




RELIC



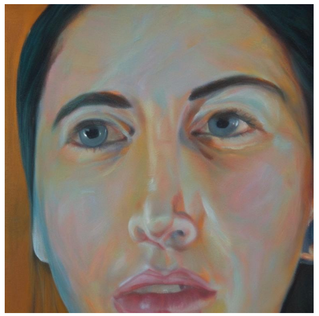

Source images

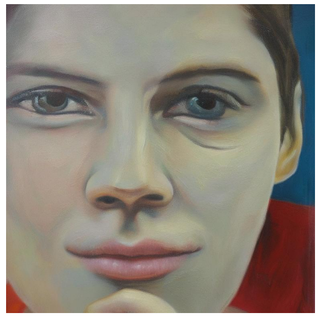
Prompt

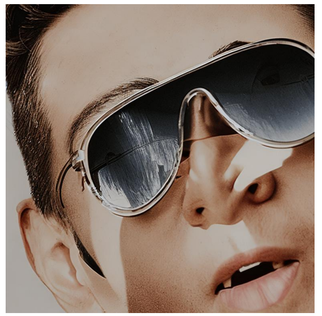
an oil painting of $S^\star$

a photo of $S^\star$ wearing sunglasses

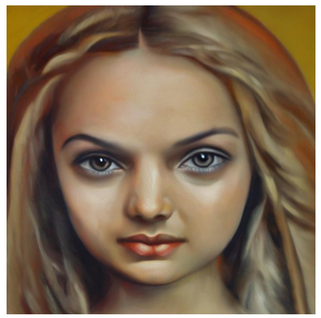
FedCPW







RELIC



Thank you