Semi-Automated Evaluation of Visual Anagrams: Assessing Perception in Diffusion Models and Humans

Flavia-Diana Macovei
University of Toronto

Motivation

- Visual anagrams: generate optical illusions using diffusion models

an oil painting of people at a campfire

an oil painting of an old man

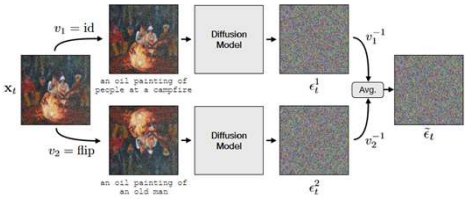
- Transformations: rotations, flips, skews, jigsaw rearrangements

Research questions:

1. How to determine the success of an illusion?
2. What are the limitations of this method?
3. Can performance be accelerated by alternating the views?

Related Work

- Diffusion models generate an image through iterative denoising towards a target
- "Visual anagrams" denoises towards two (or more) targets at each step and averages the estimate



- Scoring an illusion: view ↔ target probability

	dog	cat
id	0.95	0.05
flip	0.02	0.08

- $a(S) := \min \operatorname{diag}(S)$
- $c(S) := \frac{1}{N} \operatorname{tr}\left(\operatorname{softmax}\left(\frac{S}{\tau}\right)\right)$
- $d(S) := \operatorname{argmax}(S_{ij}) = \operatorname{argmax}(S_{kj})\ \forall i, k \in [N]$
- Human detection relies on imputed features

Future Work

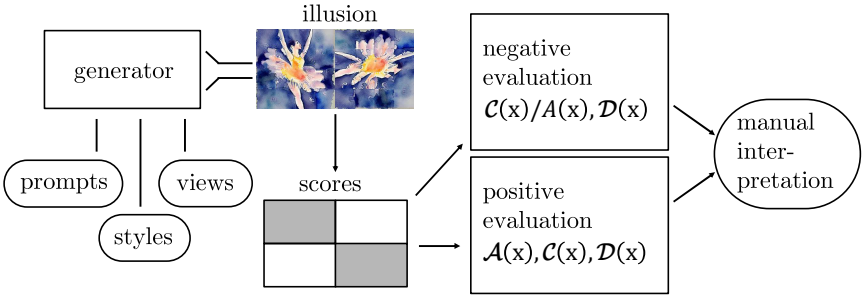
- Automated evaluation of illusions and fault detection using a large multimodal model
- In-depth investigation of techniques to reduce computational complexity

References

[1] Geng, Park, Owens, *Visual anagrams: Generating multi-view optical illusions with diffusion models*, CVPR, 2024
[2] Radford, et al., *Learning transferable visual models from natural language supervision*, ICML, 2021
[3] Schyns, Goldstone, Thibaut, *The development of features in object concepts*, Behavioral and Brain Sciences, 1998

Technique

- Automated evaluation pipeline:

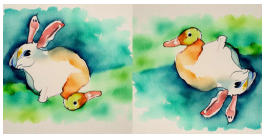


- Negative evaluation:
  - Detect common types of errors
  - Identify patterns in configurations which cause errors
  - Relate configurations to perception psychology
- Positive evaluation:
  - Implement extensions (multiple views, color channel permutation)
  - Identify configurations which lead to successful illusions
- View alternation:
  - Alternate for *n* inference steps, apply averaging for the remainder
  - Evaluate processing time and output alignment scores
  - Identify balance between computational efficiency and output quality

Experimental Results

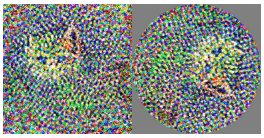
- Error types:

Independent synthesis



Dominant synthesis

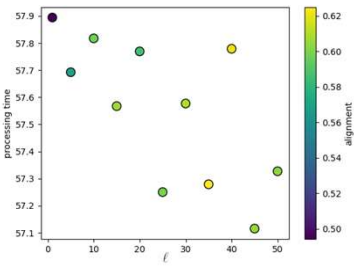
Correlated noise



- Configurations leading to independent synthesis:
  - a rabbit — a duck
  - a car — a farm
  - rotate cw — rotate 180
  - jigsaw — inner circle
- Configurations leading to dominant synthesis:
  - a cat — houseplants
  - rotate cw
  - negate — pixel permute
  - inner circle
- Correlated noise is the result of pixel misalignment of non-cardinal rotation
- Multiple views:

- Color channel permutation:

- View alternation:



- Findings diverge from original observations