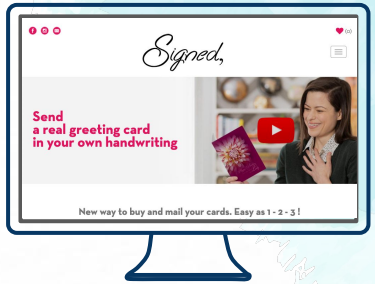
Postcards vision analysis

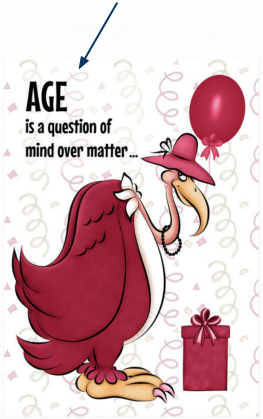
Jo Pan (tul02009@temple.edu)
CIS 5603. Artificial Intelligence

SignedCards.com

Data



Cover text



None

Birthday

None

Happy birthday

[Buzzards, vultures, birds, girly, teasing animals]

Category	Sample size	Unique Labels
Holidays	2427	29
Special Occasions	2735	29
Relationships	1847	83
Messages	2196	34
Features	6699	500

Sparse and imbalanced

Visual tasks

Classification

Holidays
Special Occasions
Relationships
Messages

Object Detection

Features
(eg: bird, balloon ...)

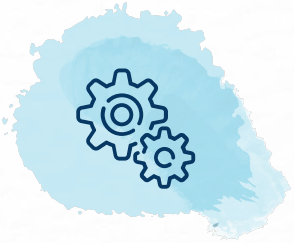


Image Retrieval

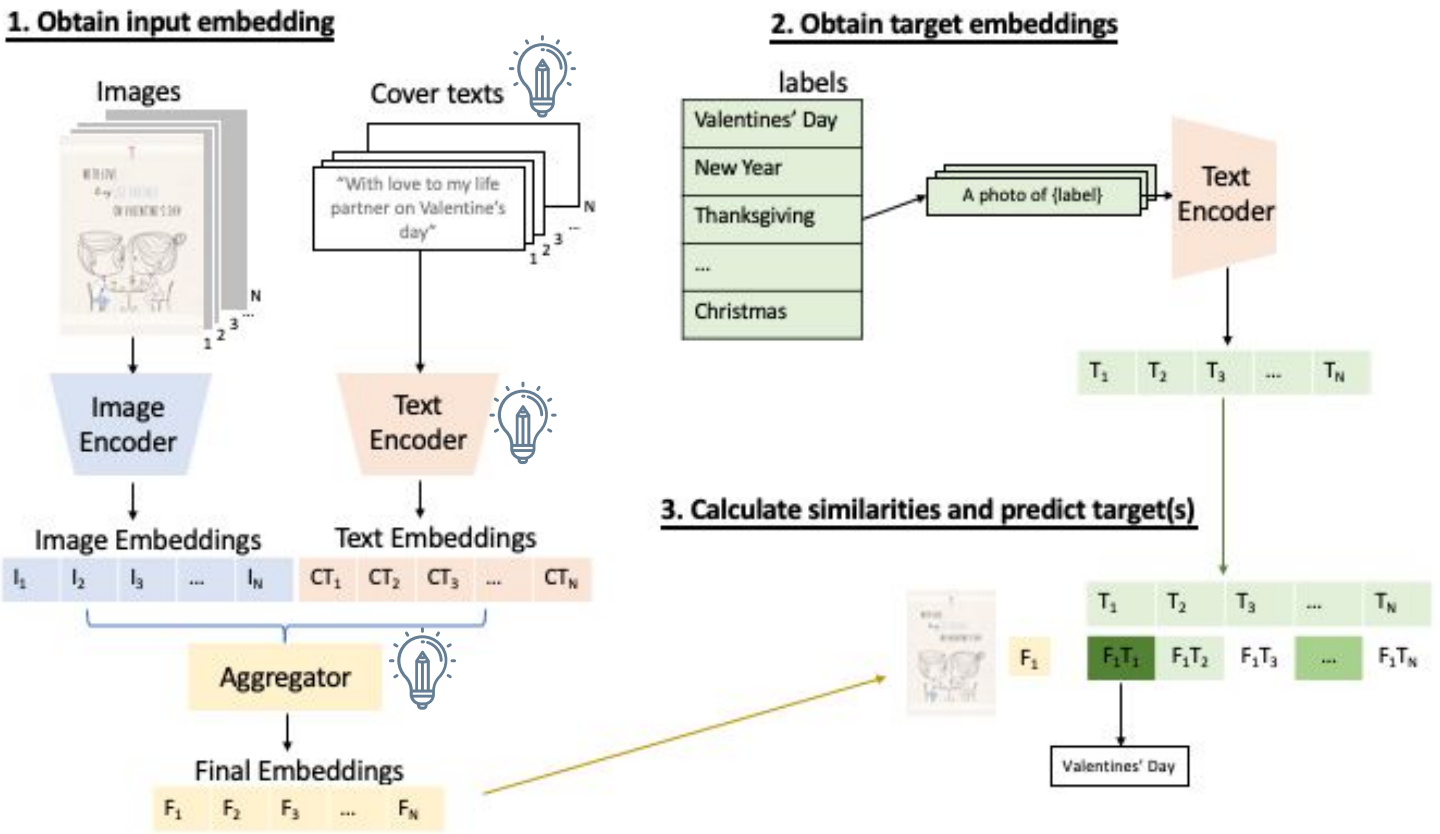
CLIP-inspired method

Improvement

Encoder choices

- CLIP (OpenAI)
- Universal Sentence Encoder (Google)

No training



Results

highlights

Classification: surprisingly good without fine-tuning

Input feature(s)	holidays		special occasions		relationships		messages	
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
image	0.632	0.950	0.531	0.877	0.338	0.565	0.278	0.509
cover text	0.775	0.891	0.595	0.942	0.449	0.732	0.336	0.508
image & text (avg.)	0.810	0.969	0.657	0.964	0.530	0.768	0.397	0.608

A trained Resnet-18 on our dataset for holiday classification:
Acc = 10%

Object detection

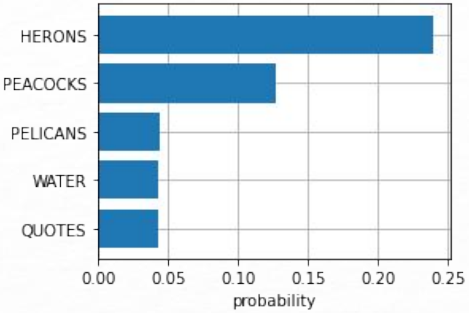
61.5%

Accuracy

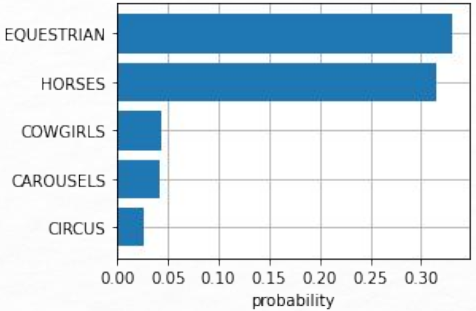
81.1%

At least 1 object correct

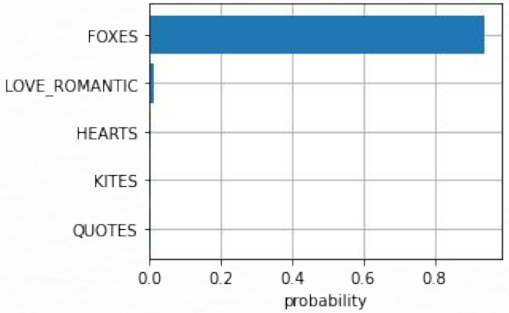
['ANIMALS', 'BIRDS']



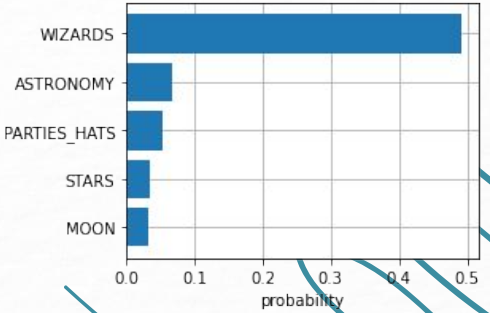
['COWGIRLS']



['ANIMALS', 'FOXES']



['WIZARDS']



Retrieve with abstract concepts

Text Queries

This is an romantic image

[1] p=0.273


[2] p=0.272


[3] p=0.270


[4] p=0.270

[5] p=0.267

This is an image with chinese

[1] p=0.290

[2] p=0.290

[3] p=0.288

[4] p=0.288


[5] p=0.287

This is an image about bon voyag

[1] p=0.287

[2] p=0.268

[3] p=0.262

[4] p=0.259

[5] p=0.253

CLIP captures large variety of concepts

t-SNE on CLIP's image embeddings