Is a “happy dog” image more “happy” than “dog”? Analyzing Adjective and Noun Visual Contributions


Dèlia Fernández

Víctor Campos

Brendan Jou

Xavier Giró-i-Nieto

Shih-Fu Chang

Columbia University in the City of New York

UPC Universitat Politècnica de Catalunya BarcelonaTech

WiML Women in Machine Learning

Introduction

1. Objective: Predict Sentiment/Emotions

Problem: Affective Gap

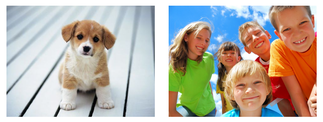
2. The Adjective-Noun Pairs (ANPs)

Solution: Mid-level Affective Representation

3. Hypothesis:

Adjective and Noun contribute differently depending on the ANP.

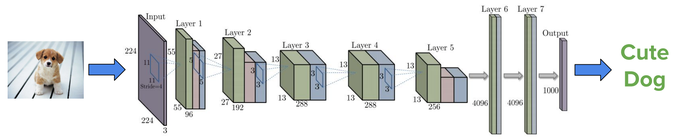
Foggy Morning — Bright Day



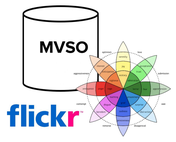
Cute Dog — Happy Kids

4. Traditional Method:

Fine-tuning Single Branch Architectures



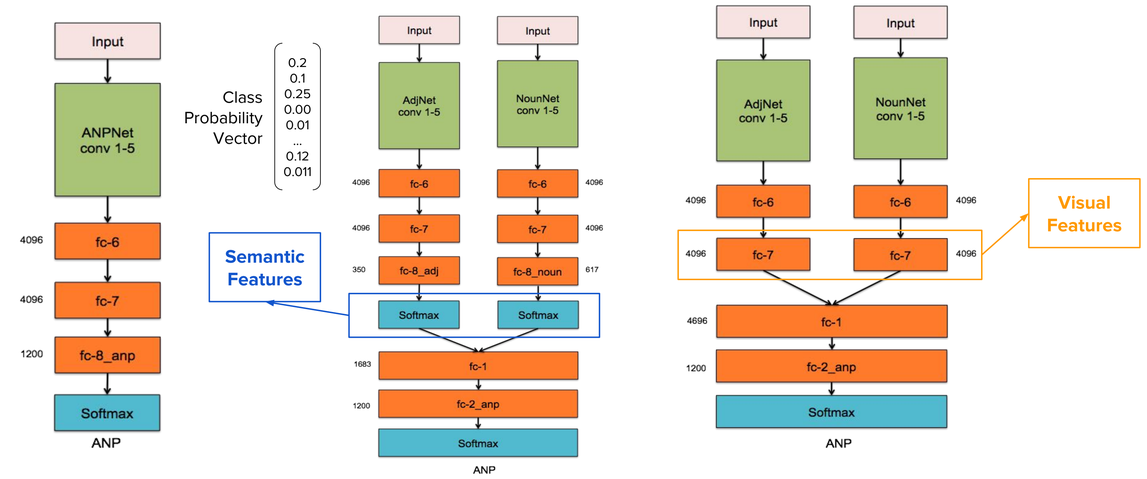
5. Multilingual Visual Sentiment Ontology (MVSO) [1]:



Subset of English-MVSO dataset:

- Tag-restricted pool of ANPs [2]
- 1,200 ANPs with more images
- All ANPs have approx. 1,000 images

Studied Architectures

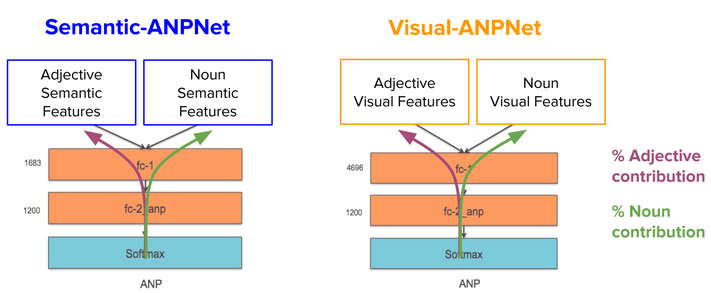


Accuracy Results

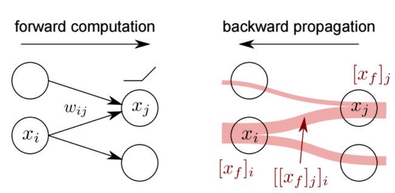
Network	#Classes	#Images	#Train	#Test	Top-1	Top-5
ANPNet (baseline)					18.03%	35.22%
Semantic-ANPNet	1,200	1,179,365	943,494	234,870	16.44%	32.68%
Visual-ANPNet					**20.02%**	**37.88%**

Backpropagate Contributions

Using Deep Taylor Decomposition [3]



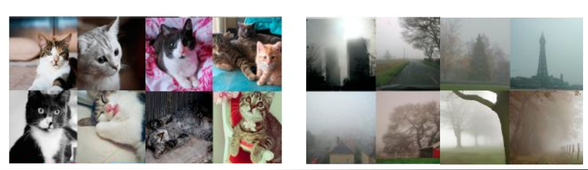
- Percentage of contribution coming from Noun-Features and from Adjective-Features
- Semantic-ANPNet: mapping features to semantic labels



Source: Montavon, Grégoire, et al. [3]

Results: % of Contribution

Noun-Oriented vs Adjective-Oriented ANPs

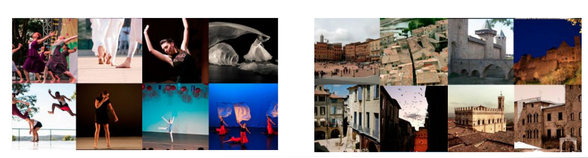


Cute Cat

% of Contribution (top-5)	Semantic	Visual
Adjectives	37.17%	30.09%
Nouns	62.82%	69.90%

Foggy Day

% of Contribution (top-5)	Semantic	Visual
Adjective	62.16%	53.42%
Nouns	37.83%	46.57%



Modern Dance

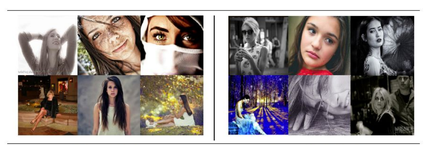
% of Contribution (top-5)	Semantic	Visual
Adjective	36.35%	41.74%
Nouns	63.64%	58.25%

Medieval City

% of Contribution (top-5)	Semantic	Visual
Adjective	52.34%	51.58%
Nouns	47.65%	48.41%

Results: Semantic Contributions

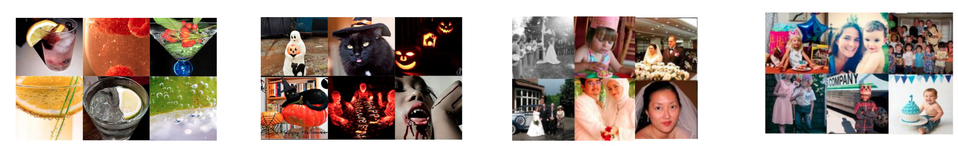
Detection of Visually Equivalent ANPs



Pretty Lady		Beautiful Lady	
top-5 adjectives	top-5 nouns	top-5 adjectives	top-5 nouns
pretty	lady	pretty	lady
beautiful	women	beautiful	women
young	model	young	model
sexy	woman	sexy	woman
happy	girl	teen	girl

Cute Cat		Cute Kitty	
top-5 adjectives	top-5 nouns	top-5 adjectives	top-5 nouns
cute	cat	cute	cat
domestic	kitty	domestic	kitty
little	kitten	little	kitten
unconditional	cats	unconditional	cats
happy	shorthair	happy	shorthair

Detection of Co-occurring Concepts



Sparkling Water		Happy Halloween		Special Day		Happy Birthday	
top-5 adjective	top-5 noun	top-5 adjective	top-5 noun	top-5 adjective	top-5 noun	top-5 adjective	top-5 noun
sparkling	water	happy	halloween	young	wedding	happy	birthday
mint	food	haunted	blood	happy	couple	traditional	cause
happy	tea	stuffed	stuff	special	bride	little	sweater
hot	pleasures	dark	comments	beautiful	day	young	kids
raw	cup	little	cat	outdoor	occasion	sweet	times

References:

[1] Jou, Brendan, et al. "Visual affect around the world: A large-scale multilingual visual sentiment ontology." ACMM, 2015.

[2] Jou, Brendan, et al. "Going Deeper for Multilingual Visual Sentiment Detection." 2016.

[3] Montavon, Grégoire, et al. "Explaining nonlinear classification decisions with deep taylor decomposition." ICML, 2015.

Gobierno de España — Ministerio de Economía y Competitividad

Generalitat de Catalunya

NVIDIA