

More cat than cute ?

Interpretable Prediction of Adjective-Noun Pairs

ACM Multimedia 2017 Workshop

Multimodal Understanding of Social, Affective and Subjective Attributes (MUSA2)



Dèlia
Fernández



Alejandro
Woodward



Víctor
Campos



Xavier
Giró-i-Nieto



Brendan
Jou



Shih-Fu
Chang

Motivation

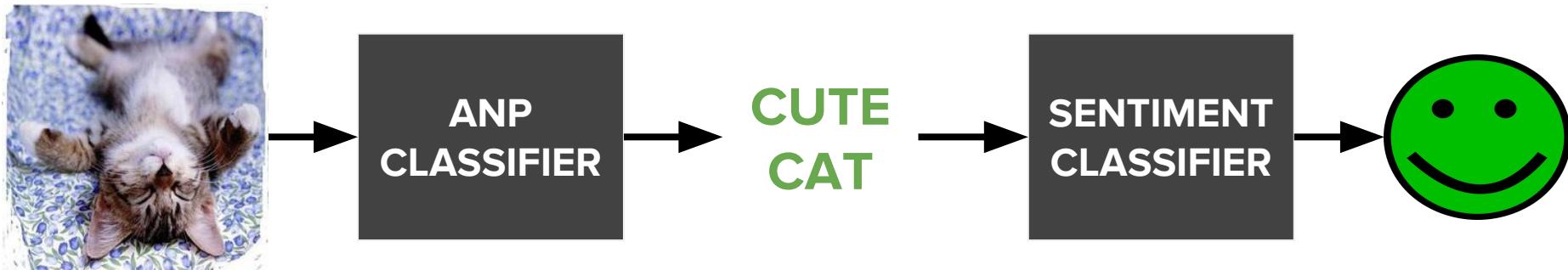
Problem: The **AFFECTIVE GAP** between low-level visual features and the emotional content of an image.



Problem: The **AFFECTIVE GAP** between low-level visual features and the emotional content of an image.

Motivation

Adjective-Noun Pairs (ANPs) provide detectable mid-level affective representations.



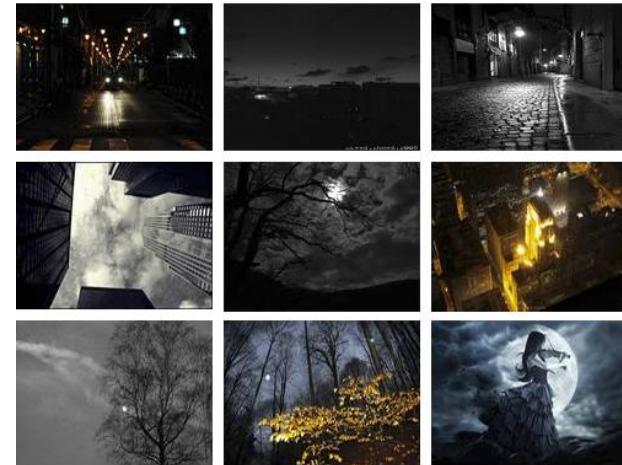
[3] Damian Borth, Rongrong Ji, Tao Chen, Thomas Breuel, and Shih-Fu Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. ACM MM 2013

Motivation

- Observation [16]: Name provides visual grounding, Adjective an affective bias.
- Our Hypothesis: For some ANPs, adjective carry most visual cues.



Cute dog



Dark night

Dataset: Subset of VSO

Query search on Flickr API:
Happy Kids



VSO Dataset [3]
2,089 ANPs

Subset



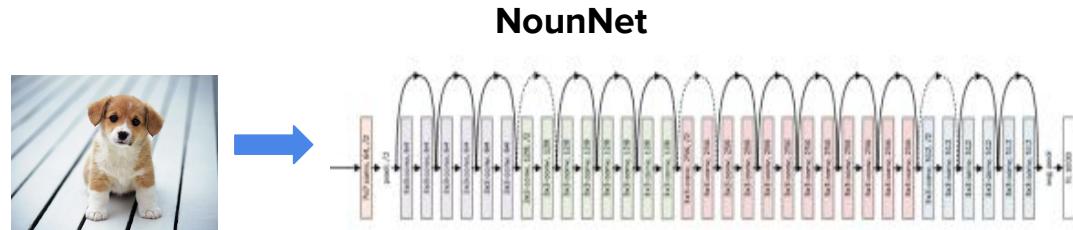
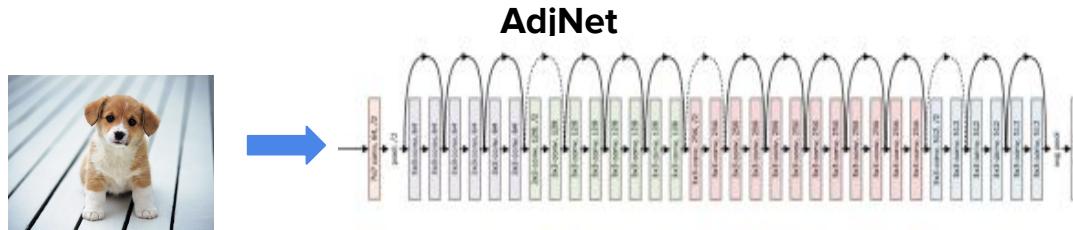
MT-VSO Dataset [14]
553 ANPs
384,258 images
(80% train - 20% tes)

[3] Damian Borth, Rongrong Ji, Tao Chen, Thomas Breuel, and Shih-Fu Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. ACM MM 2013

[14] Brendan Jou and Shih-Fu Chang. Deep Cross Residual Learning for Multi-task Visual Recognition. In ACM MM 2016

AdjNet & NounNet

Two ResNet-50 CNNs are trained independently to predict adjectives and nouns.



$$\hat{y}_{adj} = f_{adj}(x)$$

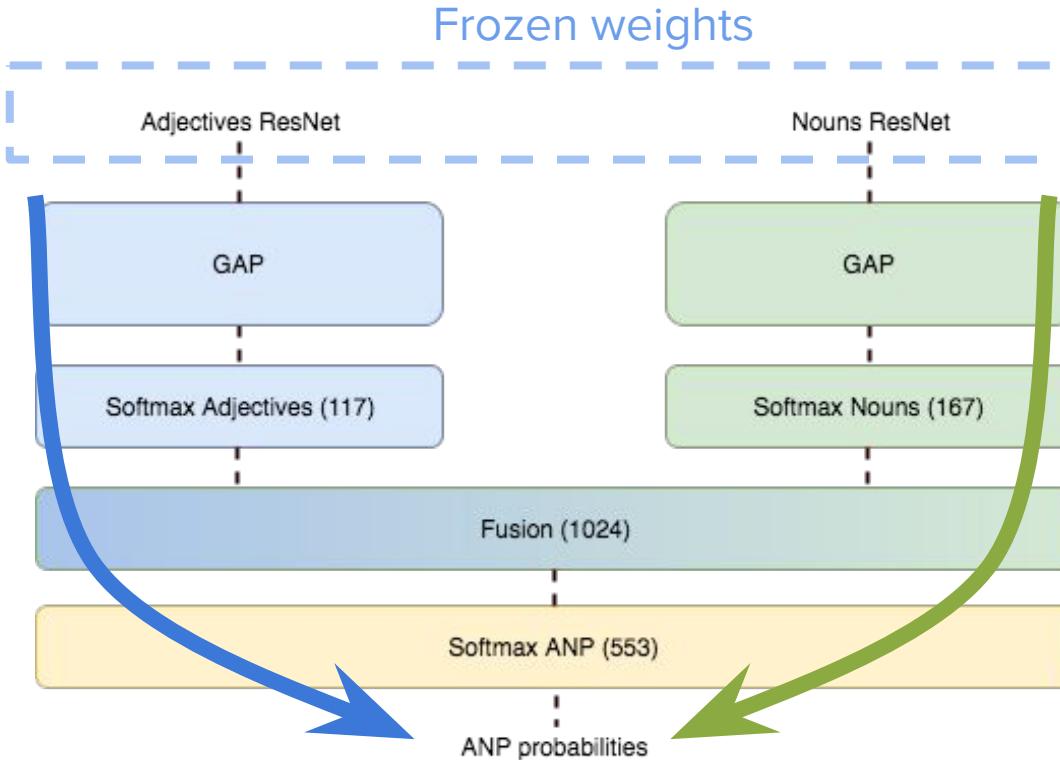
$$\hat{y}_{noun} = f_{noun}(x)$$

AdjNet & NounNet

Our implementation reproduces [14]:

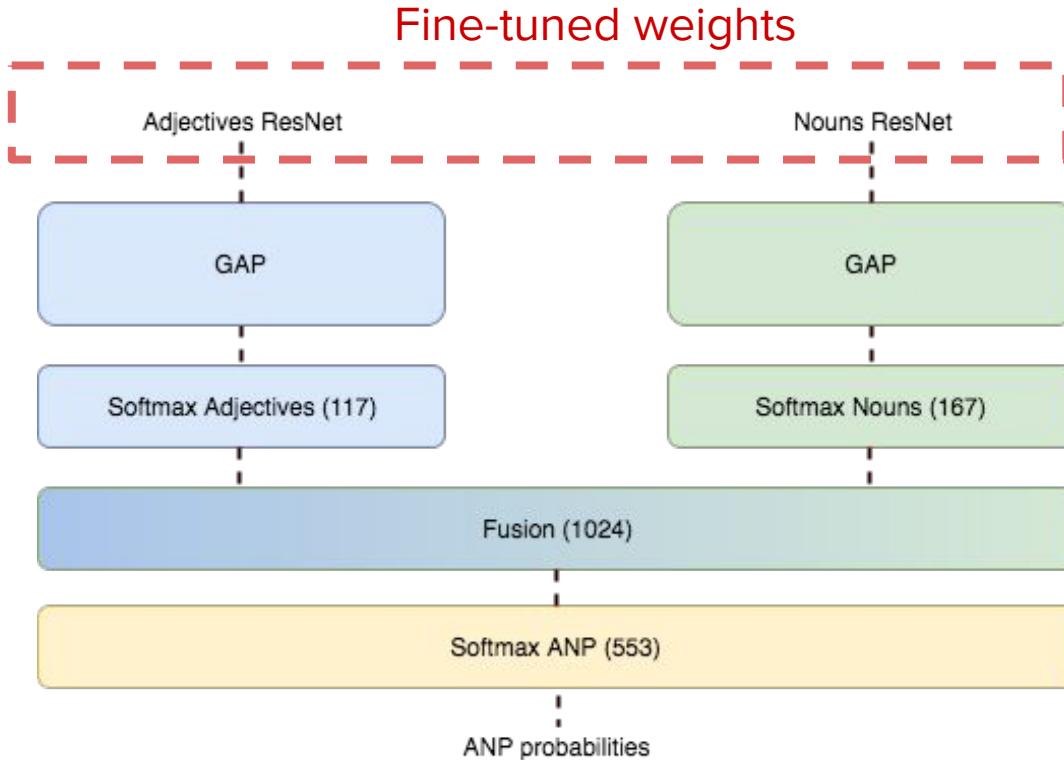
Model	Task	Classes	top-1	top-5
AdjNet [14]	Adj	117	28.45	57.87
AdjNet	Adj	117	27.70	57.00
NounNet [14]	Noun	167	41.64	69.81
NounNet	Noun	167	41.50	69.20

ANPnet



$$\hat{y}_{ANP} = g(\hat{y}_{adj}, \hat{y}_{noun})$$

Non-interpretable variation



$$\hat{y}_{ANP} = g(\hat{y}_{adj}, \hat{y}_{noun})$$

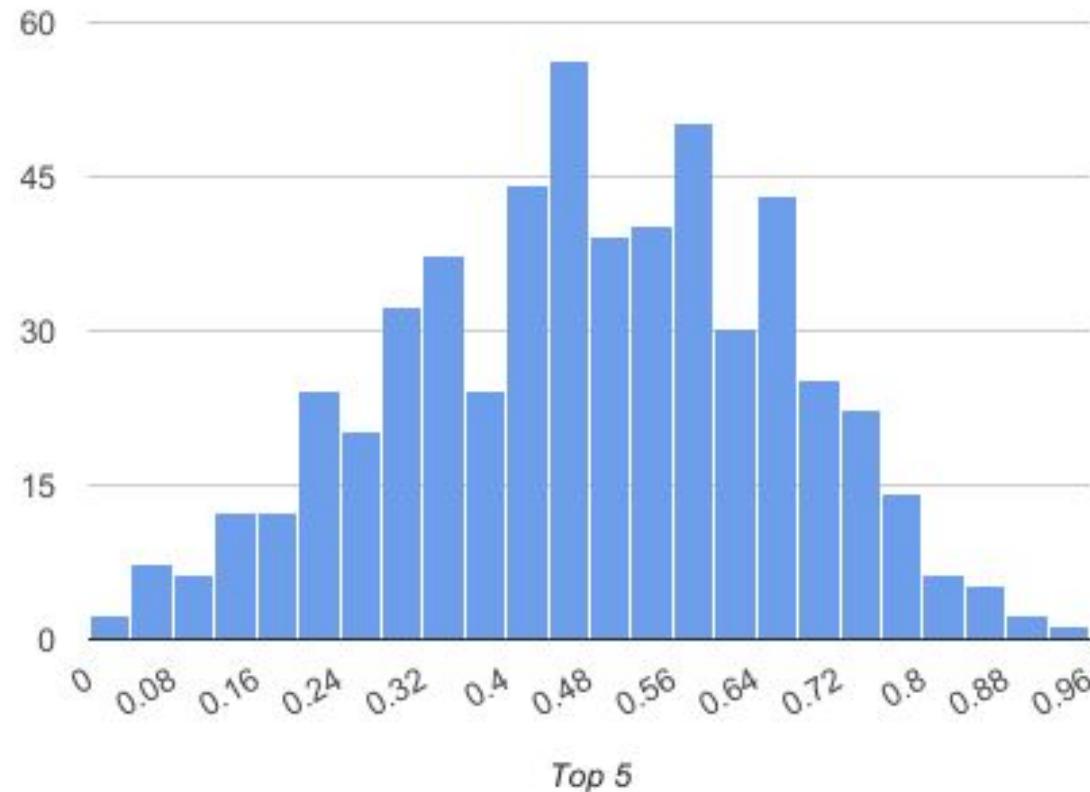
ANPnet: Top-5 Accuracy

Interpretability comes at the cost of a decrease in prediction accuracy.

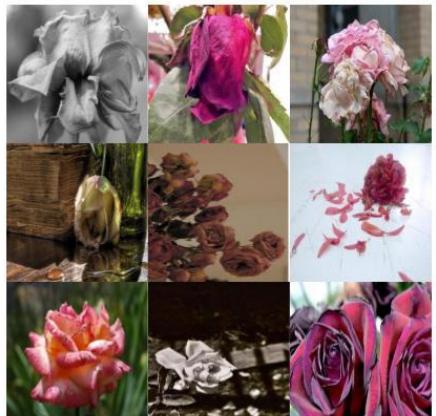
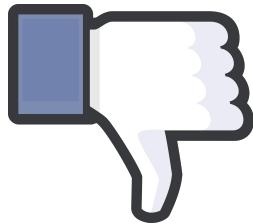
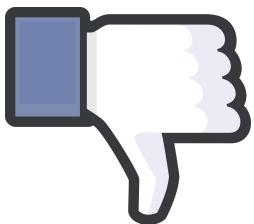
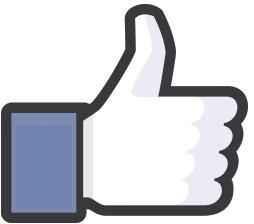
Model	Task	Classes	top-1	top-5
ResNet-50 [14]	ANP	553	22.68	47.82
ResNet-50	ANP	553	23.40	48.20
Non-Interpretable	ANP	553	21.80	46.00
ANPNet	ANP	553	20.67	43.28

ANPnet: Top-5 Accuracy

Distribution of accuracies across ANPs.



ANPnet: Top-5 Accuracy



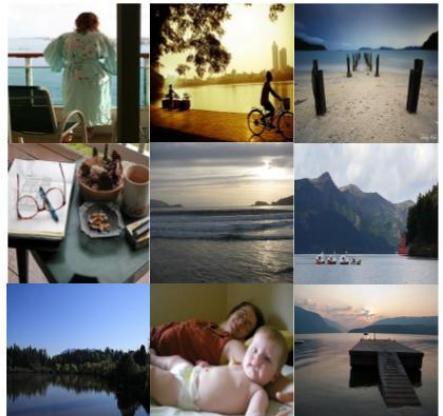
dying rose



young deer



nice scene



peaceful morning

ANPnet: Top-5 Accuracy

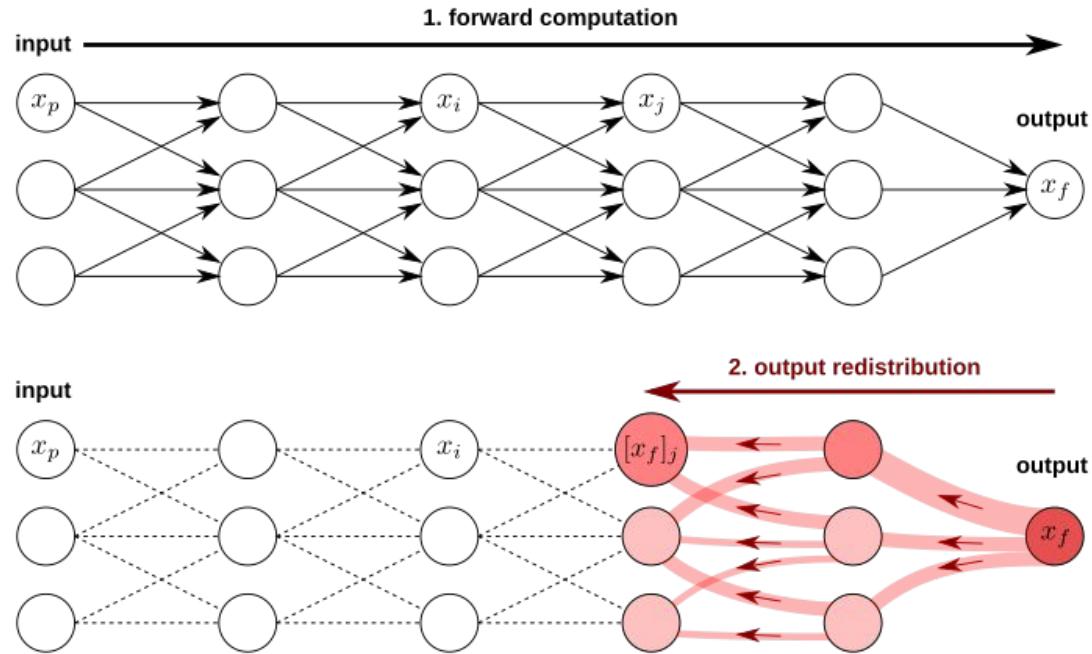
	Top-5 accuracies for best ANPs				Top-5 accuracies for worst ANPs		
	Adj	Noun	ANP		Adj	Noun	ANP
gentle river	72.22	55.97	92.45	abandoned places	84.58	33.93	8.21
tiny bathroom	70.82	88.35	91.26	beautiful landscape	60.49	42.68	7.41
young deer	52.49	89.04	89.81	beautiful earth	60.49	5.00	5.71
wild deer	64.34	89.04	86.49	charming places	4.46	33.93	5.36
misty road	79.42	80.16	86.49	bad view	38.83	67.86	5.22
dying rose	68.12	80.07	85.00	peaceful morning	26.22	53.18	4.96
icy grass	78.29	69.31	84.16	peaceful places	26.22	33.93	4.59
tiny mushrooms	70.82	82.52	84.00	nice scene	45.20	25.65	4.39
golden statue	64.56	77.12	83.61	serene scene	18.93	25.65	3.60
empty train	69.72	65.42	76.60	bright sky	52.84	67.93	3.20

Interpretability

Backpropagating contributions with Deep Taylor Decomposition [20]

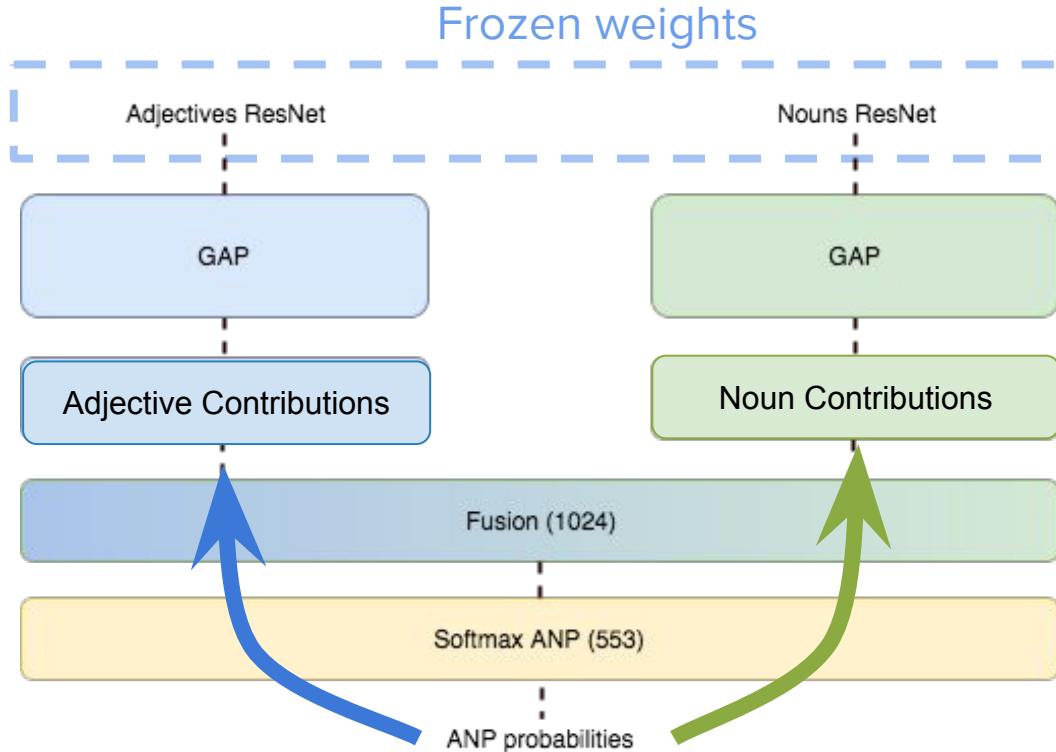
Contribution Analysis:

- Backpropagate contributions
- Adjective and Noun relevance



[20] Grégoire Montavon, Sebastian Lapuschkin, Alexander Binder, Wojciech Samek, and Klaus-Robert Müller. Explaining nonlinear classification decisions with deep taylor decomposition. Pattern Recognition (2017)

Interpretability



$$\hat{y}_{ANP} = g(\hat{y}_{adj}, \hat{y}_{noun})$$

Interpretability

$$\text{Adjective-to-Noun} \quad = \quad \frac{\text{AVERAGE[Adj contributions]}}{\text{AVERAGE[Noun contributions]}}$$

Ratio (ANR)

Interpretability

Top Adjective-oriented ANPs
(ANR > 1)

Sexy model
Misty trees
Abandoned places
Sexy body
Wild horse

Top Noun-oriented ANPs
(ANR < 1)

Innocent eyes
Incredible view
Tired eyes
Laughing baby
Chubby baby

Interpretability

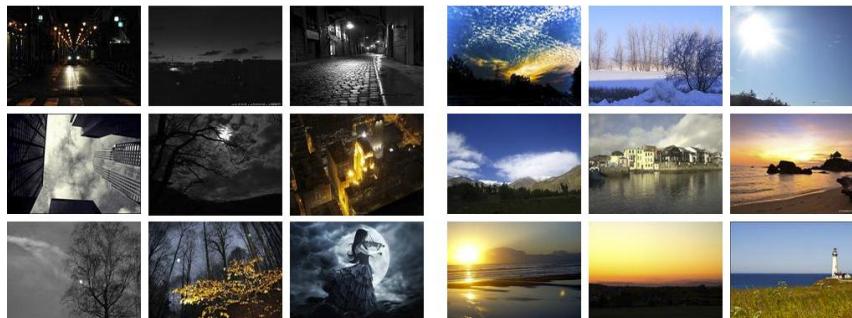
Noun oriented ANPs



a) cute dog

b) delicious cupcake

Adjective oriented ANPs



c) dark night

d) bright day

Interpretability

	Adjective-to-Noun Ratio (ANR)				All top-5 predictions ANP	Top-5 Accuracy			
	Correct predictions					Adj	Noun	ANP	
	ANP	ANP + Adj	ANP + Noun	ANP + Adj + Noun					
sexy model	1.161	1.162	1.163	1.163	1.122	76.52	62.77	59.63	
misty trees	1.139	1.140	1.138	1.139	1.146	79.42	71.74	71.88	
abandoned places	1.121	1.121	1.033	1.033	1.018	84.58	33.93	8.21	
sexy body	1.118	1.118	1.117	1.117	1.110	76.52	57.89	56.44	
wild horse	1.117	1.117	1.116	1.117	1.109	54.04	88.50	58.06	
innocent eyes	0.787	0.788	0.787	0.788	0.788	43.23	76.44	16.07	
incredible view	0.785	0.786	0.785	0.786	0.809	30.71	67.86	39.02	
tired eyes	0.776	0.778	0.776	0.788	0.784	56.13	76.44	37.50	
laughing baby	0.769	0.769	0.769	0.769	0.773	72.57	83.74	69.03	
chubby baby	0.764	0.764	0.764	0.764	0.786	48.00	83.74	45.60	



ResNet-50 was trained
with Nouns (ImageNet)

Visually equivalent ANPs

Top-5 contributing adjectives and nouns are identical.



Happy Dog

top-5 adjectives top-5 nouns

happy	dog
smiling	animals
friendly	pets
playful	grass
funny	eyes

Smiling Dog

top-5 adjectives top-5 nouns

smiling	dog
happy	eyes
friendly	pets
funny	blonde
playful	animals



Golden Autumn

top-5 adjectives top-5 nouns

golden	autumn
sunny	leaves
colorful	trees
falling	sunlight
bright	tree

Golden Leaves

top-5 adjectives top-5 nouns

golden	leaves
sunny	autumn
falling	trees
colorful	sunlight
bright	tree

Related Adjectives and Nouns

ANP tags can be extended with the most contributing adjective and nouns.



Randomly selected images

a) Elegant Wedding

b) Charming Places

top-5 adjectives top-5 nouns

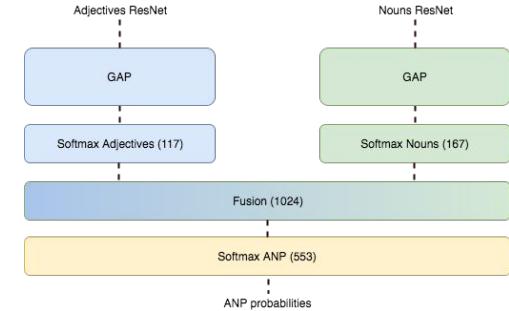
elegant	wedding
outdoor	cake
fresh	rose
tasty	dress
delicious	lady

top-5 adjectives top-5 nouns

charming	hotel
comfortable	places
excellent	house
traditional	home
expensive	food

Conclusions

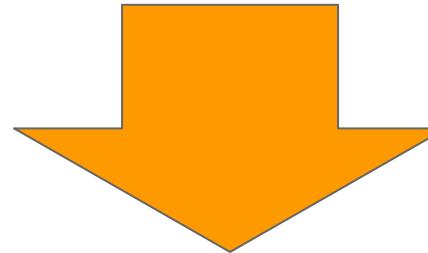
- ANPNet as an interpretable model for disentangling adjective and nouns contributions.
- Adjective-to-Nouns Ratio (ANR) as a metric to discriminate between adjective- and noun-oriented names.
- Adjective-oriented ANPs are harder to detect.
- Side products: Visually equivalent ANPs and Tag expansion.
- Interpretable model applicable to other domains.



Conclusion



$\text{ANR}(\text{Cute cat}) = 0.870$



A **cute cat** is more cat than cute



Catalonia

Spain



Directions



SAVE



NEARBY



SEND TO YOUR PHONE



SHARE



Quick facts

The Catalonia region, in northeastern Spain, is known for the lively beach resorts of Costa Brava as well as the Pyrenees Mountains. Barcelona, the regional capital, has a historic Gothic Quarter, La Rambla pedestrian mall, museums and several beaches. Antoni Gaudí's distinctive modern art and



Satellite



Map data ©2017 GeoBasis-DE/BKG (©2009), Google, Inst. Geogr. Nacional

United States

Terms

Send feedback

200 mi

DISCLAIMER: Slide included as a PERSONAL decision of the speaker.

HELP CATALONIA DEMOCRACY



Donald Tusk 
@eucopresident

Segueix

For EU nothing changes. Spain remains our only interlocutor. I hope the Spanish government favours force of argument, not argument of force.

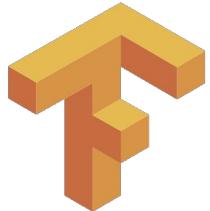
 Tradueix del anglès

16:10 - 27 oct. 2017

<https://twitter.com/eucopresident/status/923914819631271936>

DISCLAIMER: Slide included as a PERSONAL decision of the speaker.

Thank you



<http://bit.ly/acmmm17musa2>



EXCELENCIA
SEVERO
OCHOA



Generalitat
de Catalunya



Obra Social
Fundació "la Caixa"



GOBIERNO
DE ESPAÑA

MINISTERIO
DE ECONOMÍA, INDUSTRIA
Y COMPETITIVIDAD

