

# 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

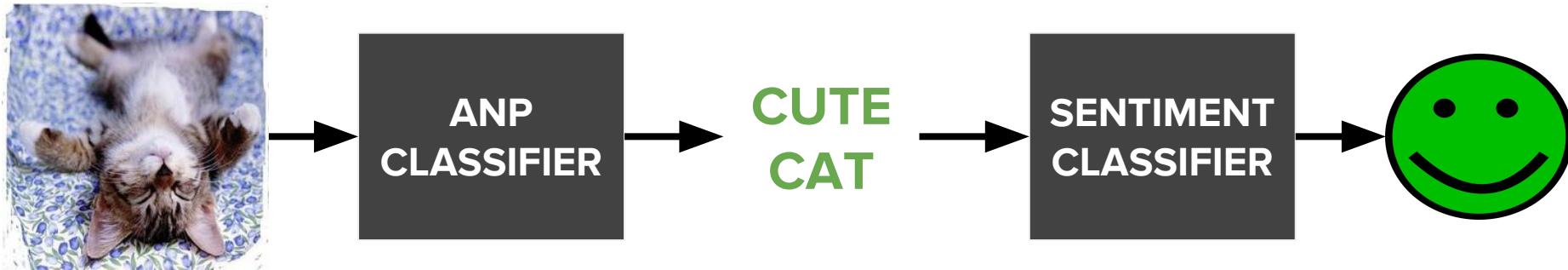
**Problem:** The **AFFECTIVE GAP** between low-level visual features and the emotional content of an image.



**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.



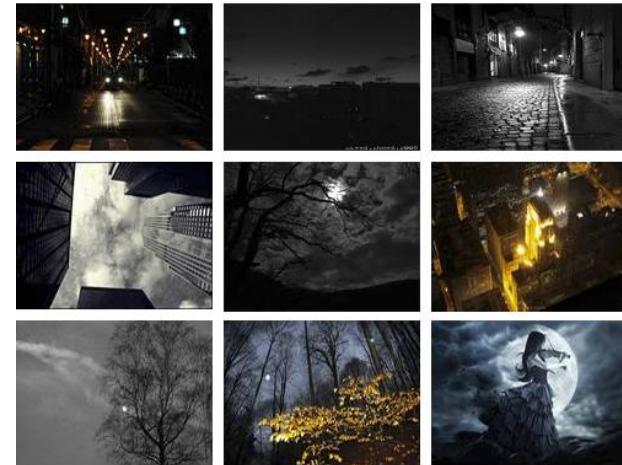
[3] Damian Borth, Rongrong Ji, Tao Chen, Thomas Breuel, and Shih-Fu Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. ACM MM 2013

# 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.

**AdjNet**



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

**NounN**



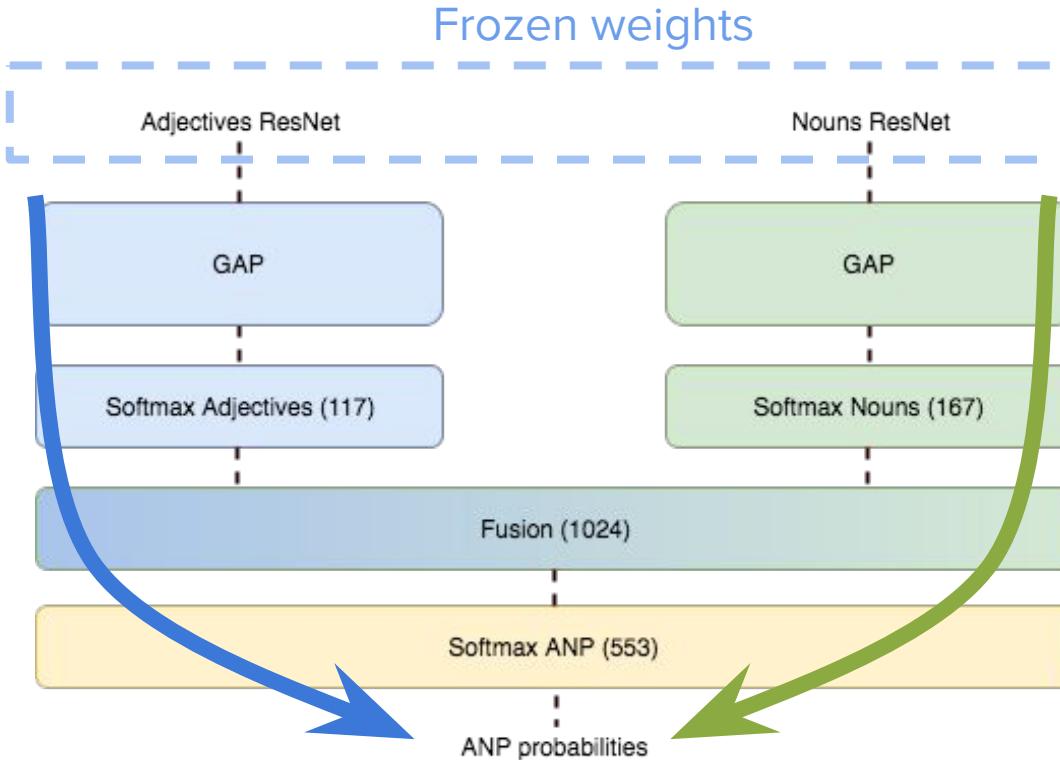
$$\hat{y}_{noun} = f_{noun}(x)$$

# AdjNet & NounNet

Our implementation reproduces [14]:

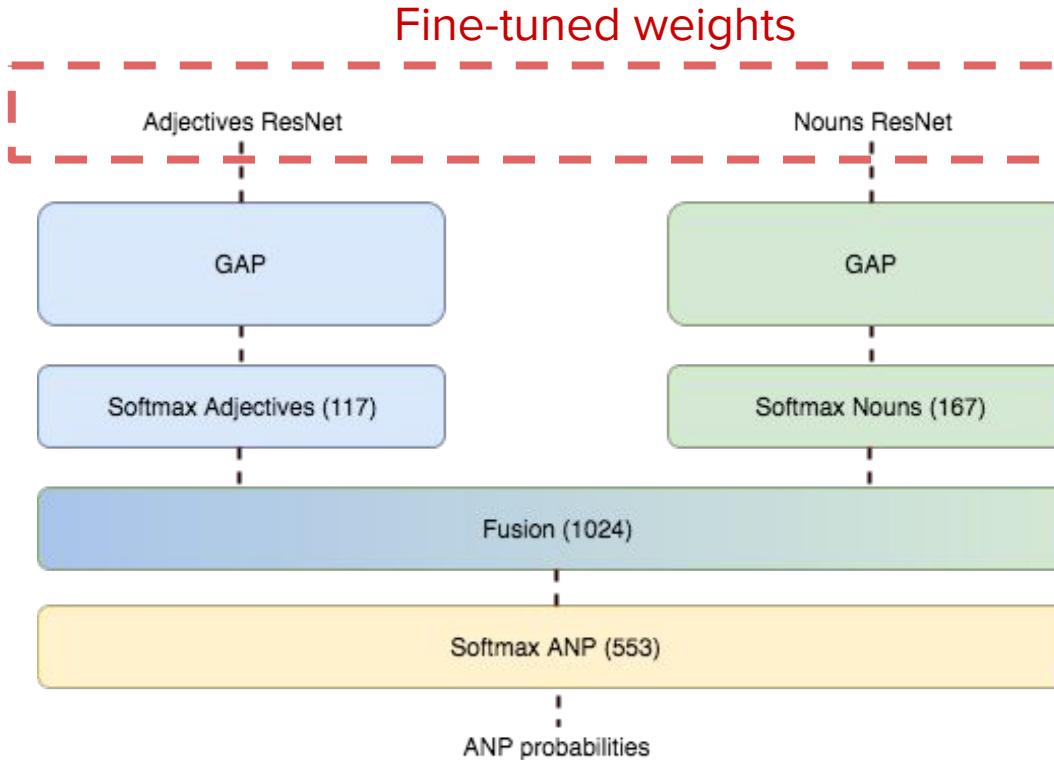
| <b>Model</b> | <b>Task</b> | <b>Classes</b> | <b>top-1</b> | <b>top-5</b> |
|--------------|-------------|----------------|--------------|--------------|
| AdjNet [14]  | Adj         | 117            | 28.45        | 57.87        |
| AdjNet       | Adj         | 117            | 27.70        | 57.00        |
| 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



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

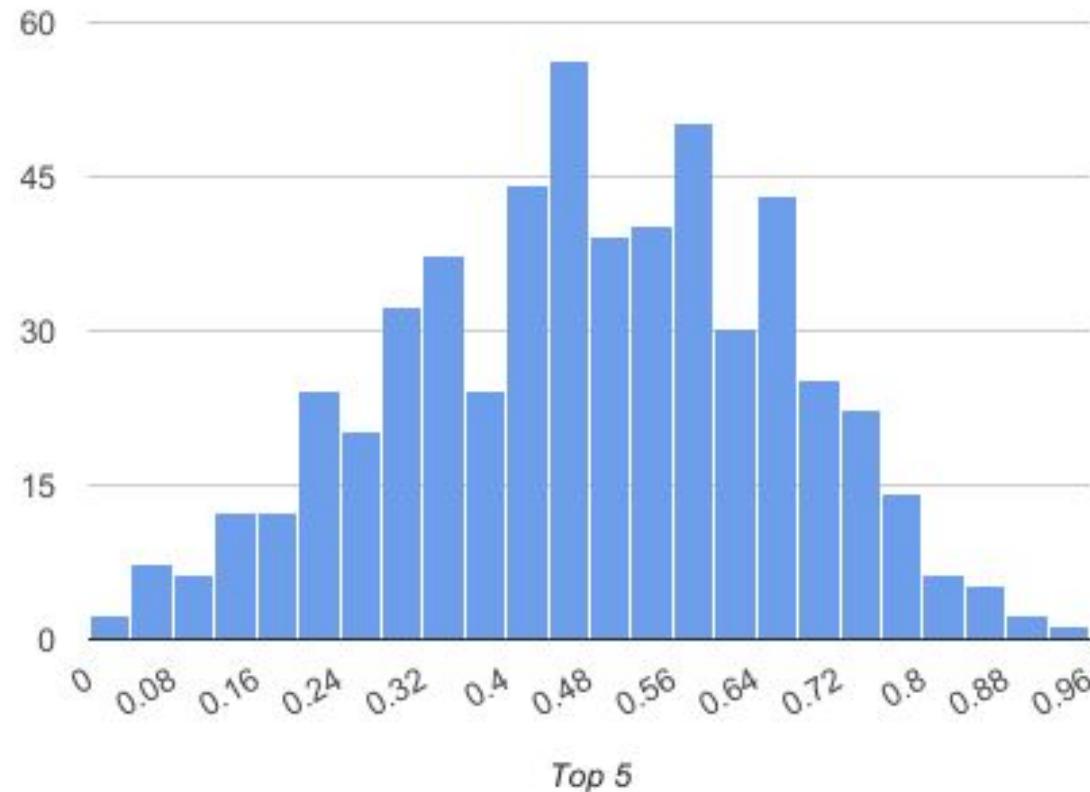
# ANPnet: Top-5 Accuracy

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

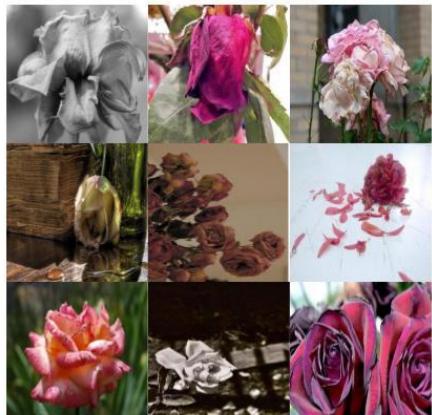
| Model             | Task | Classes | top-1 | top-5 |
|-------------------|------|---------|-------|-------|
| ResNet-50 [14]    | ANP  | 553     | 22.68 | 47.82 |
| ResNet-50         | ANP  | 553     | 23.40 | 48.20 |
| Non-Interpretable | ANP  | 553     | 21.80 | 46.00 |
| <b>ANPNet</b>     | ANP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

# ANPnet: Top-5 Accuracy

Distribution of accuracies across ANPs.



# ANPnet: Top-5 Accuracy



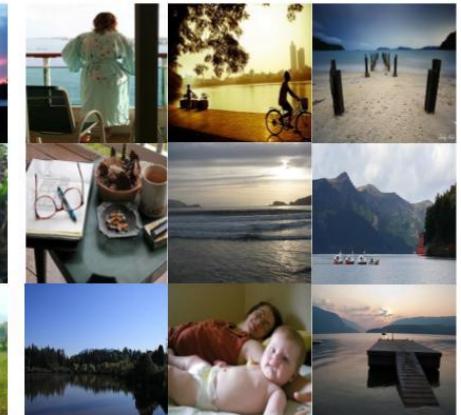
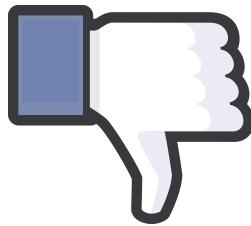
dying rose



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 |
| tiny bathroom  | 70.82                          | 88.35 | 91.26 | beautiful landscape             | 60.49 | 42.68 |
| young deer     | 52.49                          | 89.04 | 89.81 | beautiful earth                 | 60.49 | 5.00  |
| wild deer      | 64.34                          | 89.04 | 86.49 | charming places                 | 4.46  | 33.93 |
| misty road     | 79.42                          | 80.16 | 86.49 | bad view                        | 38.83 | 67.86 |
| dying rose     | 68.12                          | 80.07 | 85.00 | peaceful morning                | 26.22 | 53.18 |
| icy grass      | 78.29                          | 69.31 | 84.16 | peaceful places                 | 26.22 | 33.93 |
| tiny mushrooms | 70.82                          | 82.52 | 84.00 | nice scene                      | 45.20 | 25.65 |
| golden statue  | 64.56                          | 77.12 | 83.61 | serene scene                    | 18.93 | 25.65 |
| empty train    | 69.72                          | 65.42 | 76.60 | bright sky                      | 52.84 | 67.93 |

“Object oriented”

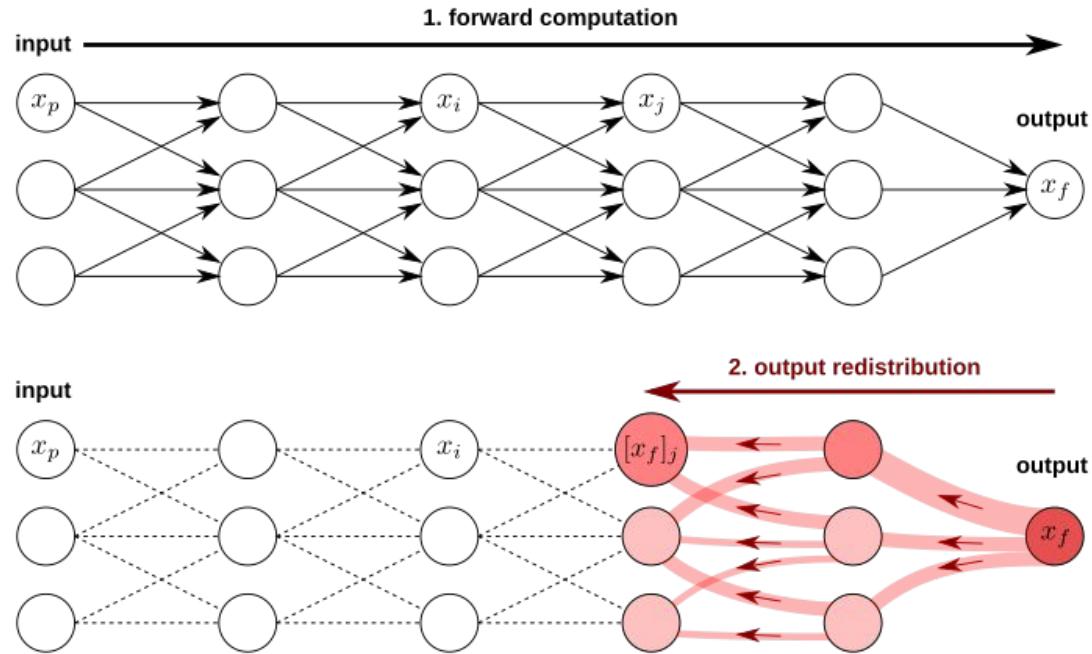
“Scene oriented”

# Interpretability

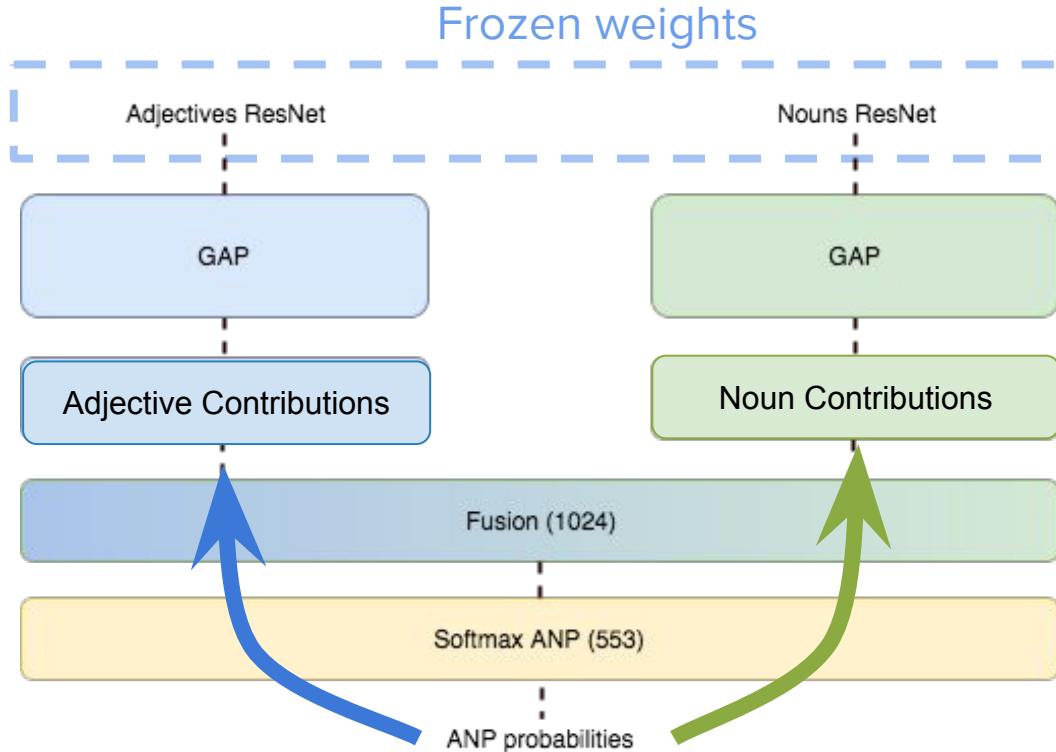
Backpropagating contributions with Deep Taylor Decomposition [20]

## Contribution Analysis:

- Backpropagate contributions
- Adjective and Noun relevance



# 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<br>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

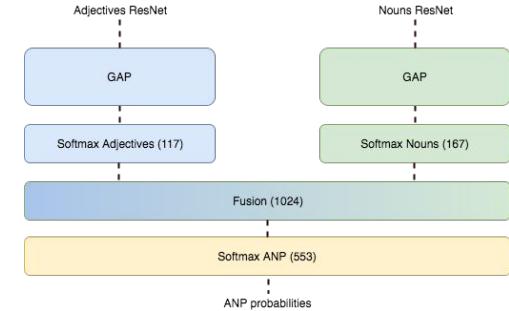
|           |         |
|-----------|---------|
| elegant   | wedding |
| outdoor   | cake    |
| fresh     | rose    |
| tasty     | dress   |
| delicious | lady    |

top-5 adjectives   top-5 nouns

|             |        |
|-------------|--------|
| charming    | hotel  |
| comfortable | places |
| excellent   | house  |
| traditional | home   |
| expensive   | food   |

# Conclusions

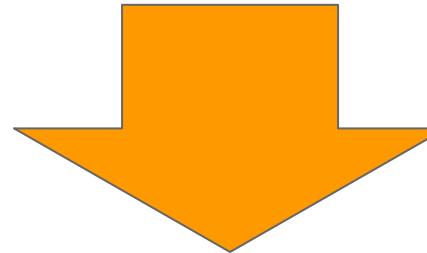
- ANPNet as an interpretable model for disentangling adjective and nouns contributions.
- Adjective-to-Nouns Ratio (ANR) as a metric to discriminate between adjective- and noun-oriented names.
- Adjective-oriented ANPs are harder to detect.
- Side products: Visually equivalent ANPs and Tag expansion.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



# Conclusion



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



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

# More Cat than Cute? Interpretable Prediction of Adjective Noun-Pairs



Délia Fernández



Alejandro Woodward



Víctor Campos



Brendan Jou



Xavier Giró



Shih-Fu Chang



COLUMBIA UNIVERSITY  
IN THE CITY OF NEW YORK



UNIVERSITAT POLITÈCNICA  
DE CATALUNYA  
BARCELONATECH



VILYNX



Biosensory  
Computing  
Multimodal  
Perception  
and Learning

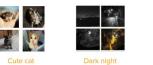


MUSA2 Workshop  
ACM Multimedia 2017

### Motivation

**1. Goal: Predict Sentiment/Emotions**  
Problem: Affective Gap  


**2. The Adjective-Noun Pairs (ANPs)**  
Solution: Mid-level Affective Representation  

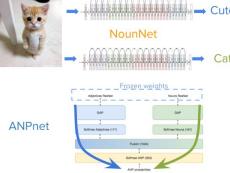

**3. Hypothesis**  
Adjective and Noun contribute differently depending on the ANP.  


**4. Dataset (VSO) [1]**  
Subset of VSO dataset:  

- Tag-restricted pool of ANPs [2]
- 553 ANPs with more images
- 384,258 images

### An Interpretable CNN Architecture

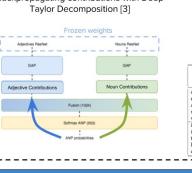
Loss in accuracy when forcing an architecture for interpretable predictions



| 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 |
| ResNet-50 [14]    | ANP  | 553     | 22.68 | 47.82 |
| ResNet-50         | ANP  | 553     | 23.40 | 48.20 |
| Non-Interpretable | ANP  | 553     | 21.80 | 43.28 |
| ANPNet            | ANP  | 553     | 20.67 | 43.28 |

### Adjective-to-Noun Ratio

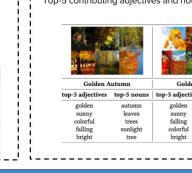
Backpropagating contributions with Deep Taylor Decomposition [3]



| Adjective-to-Noun Ratio (ANR) |           |            |                  |
|-------------------------------|-----------|------------|------------------|
| ANP                           | ANP + Adj | ANP + Noun | ANP + Adj + Noun |
| 1.00                          | 1.00      | 1.00       | 1.00             |
| 1.10                          | 1.10      | 1.10       | 1.10             |
| 1.19                          | 1.19      | 1.19       | 1.19             |
| 1.28                          | 1.28      | 1.28       | 1.28             |
| 1.38                          | 1.38      | 1.38       | 1.38             |
| 1.48                          | 1.48      | 1.48       | 1.48             |
| 1.58                          | 1.58      | 1.58       | 1.58             |
| 1.68                          | 1.68      | 1.68       | 1.68             |
| 1.78                          | 1.78      | 1.78       | 1.78             |
| 1.88                          | 1.88      | 1.88       | 1.88             |
| 1.98                          | 1.98      | 1.98       | 1.98             |
| 2.08                          | 2.08      | 2.08       | 2.08             |
| 2.18                          | 2.18      | 2.18       | 2.18             |
| 2.28                          | 2.28      | 2.28       | 2.28             |
| 2.38                          | 2.38      | 2.38       | 2.38             |
| 2.48                          | 2.48      | 2.48       | 2.48             |
| 2.58                          | 2.58      | 2.58       | 2.58             |
| 2.68                          | 2.68      | 2.68       | 2.68             |
| 2.78                          | 2.78      | 2.78       | 2.78             |
| 2.88                          | 2.88      | 2.88       | 2.88             |
| 2.98                          | 2.98      | 2.98       | 2.98             |
| 3.08                          | 3.08      | 3.08       | 3.08             |
| 3.18                          | 3.18      | 3.18       | 3.18             |
| 3.28                          | 3.28      | 3.28       | 3.28             |
| 3.38                          | 3.38      | 3.38       | 3.38             |
| 3.48                          | 3.48      | 3.48       | 3.48             |
| 3.58                          | 3.58      | 3.58       | 3.58             |
| 3.68                          | 3.68      | 3.68       | 3.68             |
| 3.78                          | 3.78      | 3.78       | 3.78             |
| 3.88                          | 3.88      | 3.88       | 3.88             |
| 3.98                          | 3.98      | 3.98       | 3.98             |
| 4.08                          | 4.08      | 4.08       | 4.08             |
| 4.18                          | 4.18      | 4.18       | 4.18             |
| 4.28                          | 4.28      | 4.28       | 4.28             |
| 4.38                          | 4.38      | 4.38       | 4.38             |
| 4.48                          | 4.48      | 4.48       | 4.48             |
| 4.58                          | 4.58      | 4.58       | 4.58             |
| 4.68                          | 4.68      | 4.68       | 4.68             |
| 4.78                          | 4.78      | 4.78       | 4.78             |
| 4.88                          | 4.88      | 4.88       | 4.88             |
| 4.98                          | 4.98      | 4.98       | 4.98             |
| 5.08                          | 5.08      | 5.08       | 5.08             |
| 5.18                          | 5.18      | 5.18       | 5.18             |
| 5.28                          | 5.28      | 5.28       | 5.28             |
| 5.38                          | 5.38      | 5.38       | 5.38             |
| 5.48                          | 5.48      | 5.48       | 5.48             |
| 5.58                          | 5.58      | 5.58       | 5.58             |
| 5.68                          | 5.68      | 5.68       | 5.68             |
| 5.78                          | 5.78      | 5.78       | 5.78             |
| 5.88                          | 5.88      | 5.88       | 5.88             |
| 5.98                          | 5.98      | 5.98       | 5.98             |
| 6.08                          | 6.08      | 6.08       | 6.08             |
| 6.18                          | 6.18      | 6.18       | 6.18             |
| 6.28                          | 6.28      | 6.28       | 6.28             |
| 6.38                          | 6.38      | 6.38       | 6.38             |
| 6.48                          | 6.48      | 6.48       | 6.48             |
| 6.58                          | 6.58      | 6.58       | 6.58             |
| 6.68                          | 6.68      | 6.68       | 6.68             |
| 6.78                          | 6.78      | 6.78       | 6.78             |
| 6.88                          | 6.88      | 6.88       | 6.88             |
| 6.98                          | 6.98      | 6.98       | 6.98             |
| 7.08                          | 7.08      | 7.08       | 7.08             |
| 7.18                          | 7.18      | 7.18       | 7.18             |
| 7.28                          | 7.28      | 7.28       | 7.28             |
| 7.38                          | 7.38      | 7.38       | 7.38             |
| 7.48                          | 7.48      | 7.48       | 7.48             |
| 7.58                          | 7.58      | 7.58       | 7.58             |
| 7.68                          | 7.68      | 7.68       | 7.68             |
| 7.78                          | 7.78      | 7.78       | 7.78             |
| 7.88                          | 7.88      | 7.88       | 7.88             |
| 7.98                          | 7.98      | 7.98       | 7.98             |
| 8.08                          | 8.08      | 8.08       | 8.08             |
| 8.18                          | 8.18      | 8.18       | 8.18             |
| 8.28                          | 8.28      | 8.28       | 8.28             |
| 8.38                          | 8.38      | 8.38       | 8.38             |
| 8.48                          | 8.48      | 8.48       | 8.48             |
| 8.58                          | 8.58      | 8.58       | 8.58             |
| 8.68                          | 8.68      | 8.68       | 8.68             |
| 8.78                          | 8.78      | 8.78       | 8.78             |
| 8.88                          | 8.88      | 8.88       | 8.88             |
| 8.98                          | 8.98      | 8.98       | 8.98             |
| 9.08                          | 9.08      | 9.08       | 9.08             |
| 9.18                          | 9.18      | 9.18       | 9.18             |
| 9.28                          | 9.28      | 9.28       | 9.28             |
| 9.38                          | 9.38      | 9.38       | 9.38             |
| 9.48                          | 9.48      | 9.48       | 9.48             |
| 9.58                          | 9.58      | 9.58       | 9.58             |
| 9.68                          | 9.68      | 9.68       | 9.68             |
| 9.78                          | 9.78      | 9.78       | 9.78             |
| 9.88                          | 9.88      | 9.88       | 9.88             |
| 9.98                          | 9.98      | 9.98       | 9.98             |

### Equivalent ANPs

Top-5 contributing adjectives and nouns are identical.



| top-5 adjectives | top-5 nouns | top-5 adjectives | top-5 nouns |
|------------------|-------------|------------------|-------------|
| golden           | autumn      | golden           | autumn      |
| sunny            | leaves      | sunny            | leaves      |
| outdoor          | trees       | outdoor          | trees       |
| fresh            | falling     | fresh            | falling     |
| bright           | height      | bright           | height      |

### Related ANPs

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



| a) elegant Wedding | b) charming Places |
|--------------------|--------------------|
| top-5 adjectives   | top-5 nouns        |

Model and source code:  
<http://bit.ly/musa2>

References:

- [1] Damian Borth et al. "Large-scale visual sentiment ontology and detectors using adjective noun pairs" ACM MM 2013
- [2] M. Montavon, G. Montavon, and K.-R. Müller. "Explaining nonlinear classification decisions with deep taylor decomposition" ICML 2015

Generalitat  
de Catalunya

nVIDIA

